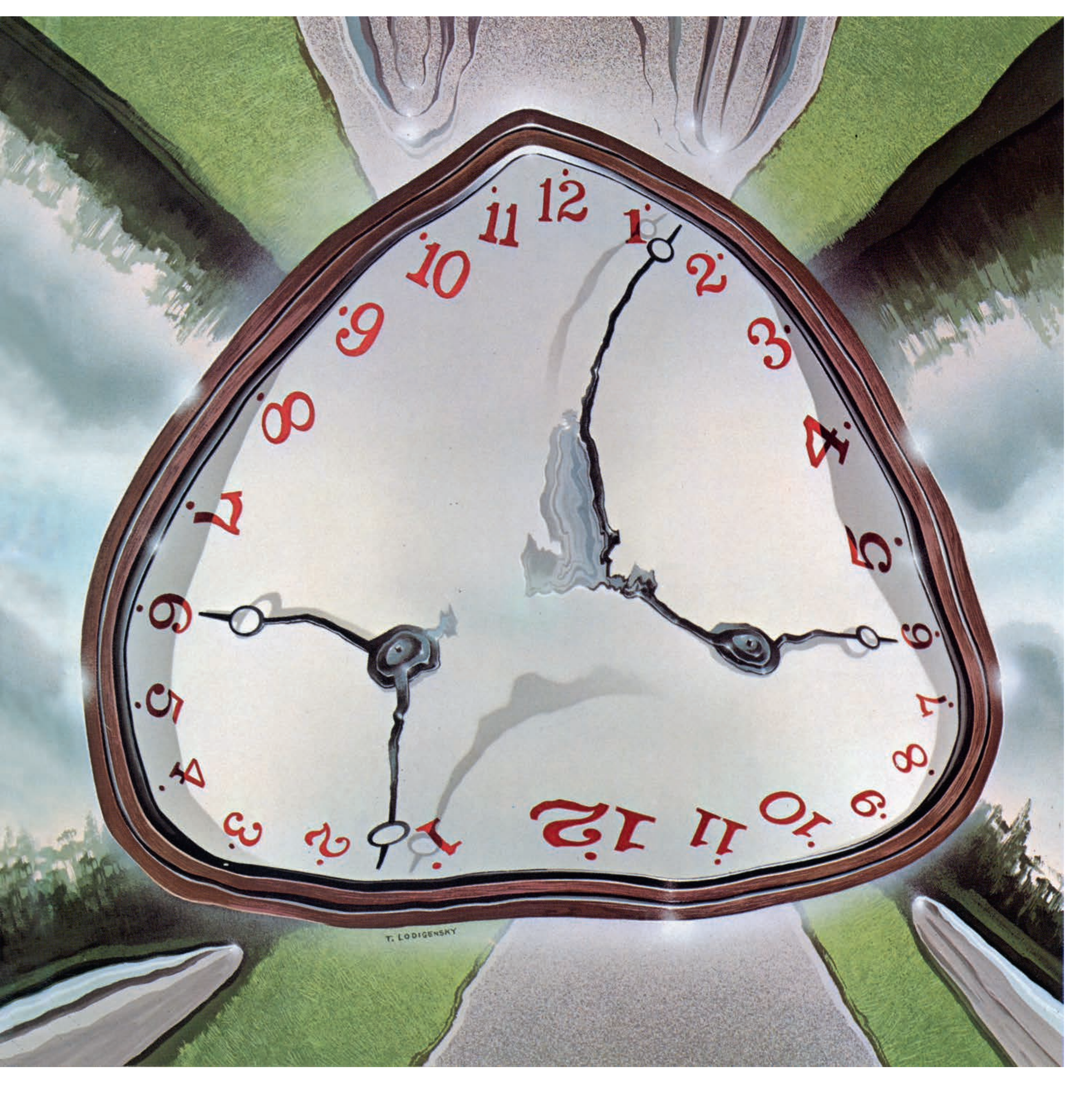
INVESTIGACION Y CIENCIA

Edición en español de SCIENTIFIC AMERICAN



IMAGENES Y ESPEJOS

Febrero 1981

250 PTAS.

Los espacios en gris
corresponden a publicidad
en la edición impresa

Número 53
Febrero 1981

INVESTIGACION Y
CIENCIA *Edición en español de* SCIENTIFIC AMERICAN

SCIENTIFIC AMERICAN

COMITE DE REDACCION: Gerard Piel (Presidente), Dennis Flanagan, Brian P. Hayes, Philip Morrison, Francis Bello, Michael Feirtag, Judith Friedman, Paul W. Hoffman, Jonathan B. Piel, John Purcell, James T. Rogers, Armand Schwab, Jr., Joseph Wisnovsky

DIRECCION EDITORIAL: Dennis Flanagan
DIRECCION ARTISTICA: Samuel L. Howard
PRODUCCION: Richard Sasso
DIRECTOR GENERAL: George S. Conn

INVESTIGACION Y CIENCIA

DIRECTOR: Francisco Gracia Guillén
REDACCION: José María Valderas Gallardo (Redactor Jefe)
Carlos Oppenheimer
Encarna de Blas (Secretaria de Redacción)
PRODUCCION: César Redondo Zayas
PROMOCION Y PUBLICIDAD: Elena Sánchez-Fabres
PROMOCION EXTERIOR: Pedro Clotas Cierco
EDITA: Prensa Científica S.A.
Calabria, 235-239
Barcelona-29 (ESPAÑA)

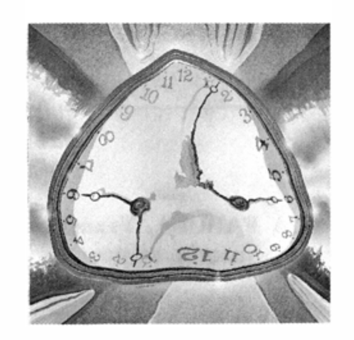
Colaboradores de este número:

Asesoramiento y traducción:

Julio Samsó: *Galileo observa Neptuno;* Ramón Pascual: *Asimetría cósmica entre materia y antimateria;* Enrique Cerdá: *Plásmidos;* Jaime Truyols: *El hombre primitivo en el Asia central soviética;* Manuel Valdés: *La conducta adictiva;* Joandomènec Ros: *El sistema de huida de la cucaracha;* Amando García: *Formación de imágenes en los espejos;* Luis Bou: *Juegos matemáticos;* J. Vilardell: *Taller y laboratorio.*

Libros:

Juan José Acero, F. Javier García-Conde y Luis Alonso



LA PORTADA

La pintura de la portada representa la imagen de la esfera de un reloj tal como se refleja en un espejo con tres lóbulos, o colinas, y tres valles. La imagen que se forma es única y continua, pero la esfera del reloj resulta ser doble, es decir, está constituida por un total de 24 horas. Si observáramos esta imagen durante 12 horas, veríamos que, mientras la saeta horaria del reloj real realiza una vuelta completa, cada una de las dos saetas horarias del reloj imagen efectúa media vuelta. Unicamente al cabo de 24 horas cada saeta habría dado una vuelta completa. Las propiedades reflectoras de las diferentes clases de superficies se pueden analizar en términos de un pequeño número de espejos curvados básicos y las fronteras que los unen (véase "La formación de imágenes en los espejos", por David Emil Thomas en este número). Las formas que se observan alrededor del reloj representan la imagen distorsionada del jardín del artista, en el que se colocaron el reloj y el espejo que utilizó de modelo para pintar el cuadro.

Suscripciones:
Prensa Científica, S. A.
Calabria, 235-239
Barcelona-29 (España)
Teléfono 322 05 51 ext. 41

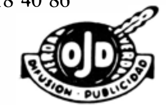
Condiciones de suscripción:
España:
Un año (12 números): 2.750 pesetas
Extranjero:
Un año (12 números): 43 U.S.$
Ejemplar atrasado ordinario:
280 pesetas
Ejemplar atrasado extraordinario:
420 pesetas

Distribución para España
Distribuciones de Enlace, S. A.
Ausias March, 49, Barcelona-10

Distribución para los restantes países:
Editorial Labor, S. A.
Calabria, 235-239 - Barcelona-29

Publicidad:
Madrid:
Gustavo Martínez Ovin
Avda. de Moratalaz, 137, Madrid-30
Tel. 430 84 81
Cataluña:
Miguel Munill
Balmes, 191, 2.°, 2.ª, Barcelona-6
Tels. 218 44 45 y 218 40 86

Controlado por O.J.D.



PROCEDENCIA DE LAS ILUSTRACIONES

Pintura de la portada de Ted Lodigensky

Página	Fuente	Página	Fuente
6-19	Francisco Rodríguez Adrados	61	Richard S. Davis, Bryn Mawr College (*arriba*); Vadim A. Ranov, Instituto de Historia de la Academia de Ciencias de Tadjikistán (*abajo*)
23-26	Instituto y Museo de la Historia de la Ciencia de Florencia		
27-28	Alan D. Iselin		
29	Instituto y Museo de la Historia de la Ciencia de Florencia	62	National Aeronautics and Space Administration, U.S. Geological Survey, EROS Data Center
30	Alan D. Iselin		
33	Brookhaven National Laboratory	63	Andrew Tomko
		64-65	Patricia J. Wynne
34-40	Gabor Kiss	66	Andrew Tomko
46	Richard P. Novick, Public Health Research Institute of the City of New York	69	Dover Publications, Inc.
		70-73	Ilil Arbel
		77-85	Tom Prentiss
		87	Ralph Morse
48	Richard P. Novick, Public Health Research Institute of the City of New York (*izquierda*); Ilil Arbel (*derecha*)	88-91	Dan Todd
		92-95	Ralph Morse
		96	Dan Todd
		97-98	Ralph Morse
		99	Dan Todd (*arriba, izquierda*) Ralph Morse (*arriba, derecha*)
49	Daniel A. Portnoy y Stanley Falkow, Universidad de Washington en Seattle	100	Ralph Morse
		102-106	Ilil Arbel
		108	Gera Dillon
50-59	Bunji Tagawa	110-113	Michael Goodman

ISSN 0210-136X
Dep. legal: B. 38.999-76
Fotocomposición Tecfa
Guipúzcoa, 36 (local 1) Barcelona-20
Fotocromos reproducidos por GINSA, S. A.
Imprime GRAFESA
Gráfica Elzeviriana, S. A.
Nápoles, 249 - Tel. 207 40 11
Barcelona-13
Printed in Spain - Impreso en España

Los autores

FRANCISCO RODRIGUEZ ADRADOS ("La fábula"), nacido en Salamanca en 1922, es catedrático de filología griega de la Universidad Complutense de Madrid desde 1952, después de haberlo sido de la de Barcelona en 1951. Ha sido presidente de la Sociedad Española de Estudios Clásicos y de la de Lingüística. Dirige las revistas "Emerita" "Española de Lingüística", así como el Gran Diccionario Griego-Español, cuyo primer volumen acaba de aparecer, y la Colección Hispánica de Autores Griegos y Latinos. Ha publicado más de veinte libros sobre literatura y filología griega antigua, lingüística indoeuropea y general, algunos de ellos traducidos al inglés y al alemán. Una de sus especialidades es el estudio de la fábula, que comenzó con sus "Estudios sobre el léxico de las fábulas esópicas" (1948) y al que dedica una amplia "Historia de la fábula greco-latina" en curso de publicación.

STILLMAN DRAKE y CHARLES T. KOWAL ("Galileo observa Neptuno") son, respectivamente, profesor jubilado de historia de la ciencia en la Universidad de Toronto y miembro del departamento de astronomía del Instituto de Tecnología de California. Drake estudió en la Universidad de California en Berkeley. En 1967 se incorporó al Instituto de Historia y Filosofía de la Ciencia de Toronto. Se jubiló en 1978. Kowal se licenció en astronomía por la Universidad de California del Sur, en 1963, y, ese mismo año, se trasladó al Instituto de Tecnología de California. Antes de esta fecha, no obstante, ya había trabajado como ayudante de investigación en los observatorios de Monte Wilson y Monte Palomar. Entre sus éxitos más notables en el campo de la observación se cuentan los descubrimientos de 81 supernovas y, de manera especial, la brillantísima supernova de 1972. En 1974 descubrió el decimotercer satélite de Júpiter y, en 1977, un planetoide llamado Quirón, situado entre las órbitas de Saturno y Urano. En 1979, la Academia Nacional de Ciencias le concedió el premio James Craig Watson.

FRANK WILCZEK ("La asimetría cósmica entre materia y antimateria") es profesor de física de la Universidad de Princeton. Por comisión de servicio se halla hoy en el Instituto de Física Teórica de la Universidad de California en Santa Bárbara. Se licenció en matemáticas por la Universidad de Chicago en 1970, doctorándose en física por la de Princeton en 1974, en cuyo claustro docente ingresó. Desde 1976 hasta 1978 fue becario de la Fundación Alfred P. Sloan y, en 1977 y 1978, del Institute for Advanced Study de Princeton.

RICHARD P. NOVICK ("Plásmidos") dirige el departamento de biología de plásmidos del Instituto de Investigación de la Salud Pública de la ciudad de Nueva York. Licenciado por la Universidad de Yale en 1954, se doctoró en medicina por la de Nueva York en 1959. Durante sus estudios dedicó un año a trabajar, con Werner K. Maas, de este último centro superior, sobre regulación genética en bacterias, iniciándose así en la genética molecular. Pasó un año en Yale y, dos, con Martin Pollock en el Instituto Nacional británico de Investigación Médica. Allí descubrió la existencia de plásmidos en la bacteria *Staphylococcus aureus,* de donde parte su interés por esos fragmentos extracromosómicos de ADN. Indeciso sobre si dedicarse a la investigación básica o ejercer su profesión médica, trabajo durante un año en la Universidad de Vanderbilt como médico residente. En 1965 se sumó a la plantilla del Instituto de Investigación de la Salud Pública.

RICHARD S. DAVIS, VADIM A. RANOV y ANDREY E. DODONOV ("El hombre primitivo en el Asia central soviética") son, respectivamente, profesor del departamento de antropología de la Facultad Bryn Mawr, jefe de la sección de arqueología y numismática del Instituto de Arqueología de la Academia de Ciencias de la R.S.S. del Tadjikistán y miembro del Instituto de Geología de la Academia de Ciencias de la Unión Soviética. En 1977 Davis pasó seis meses en Tadjikistán, en el Asia central soviética, en intercambio científico de programas entre la Academia Nacional de Ciencias de los Estados Unidos y la Academia de Ciencias de la URSS. A lo largo de ese tiempo trabajó con Ranov y Dodonov en arqueología del Paleolítico. Volvió a reanudar la colaboración con Ranov cuando éste visitó los Estados Unidos en un programa de intercambio entre el International Research and Exchanges Board (IREX) y la Academia soviética. Davis se graduó por el Columbia College en 1966 y se doctoró por la Universidad de Columbia en 1974. Ranov ha participado desde 1953 en numerosas excavaciones y proyectos del servicio en el Asia central soviética, desde los llanos desérticos inmediatos al Amu Darya hasta los altos valles del Pamir oriental. Dodonov es un geólogo unido en varias ocasiones a Ranov en la solución de problemas de interpretación de depósitos que encierran utillaje.

VINCENT P. DOLE ("La conducta adictiva") es profesor de la Universidad Rockefeller y médico de su hospital clínico. Se licenció en matemáticas por la Universidad de Stanford en 1934 y se doctoró por la Facultad de Medicina de Harvard en 1939. En 1941 se incorporó a la plantilla del Instituto Rockefeller de Investigación Médica. En los últimos años se ha interesado prioritariamente por el problema del abuso de narcóticos y por los aspectos científicos y clínicos del tratamiento con metadona, que es el tema que desarrolla en este artículo. Anteriormente le había preocupado el metabolismo de los ácidos grasos libres y del tejido adiposo. Dole escribe: "El único asunto importante no mencionado en mi curriculum es mi gran deuda con Marie Nyswander, mi íntima colaboradora profesional y esposa. Ella me enseñó a ver a los adictos como pacientes y no sólo como problemas".

JEFFREY M. CAMHI ("El sistema de huida de la cucaracha") enseña biología en la Universidad de Cornell. Se formó en la Universidad de Tufts en 1963, aunque obtuvo su doctorado por la de Harvard en 1968. Se incorporó a la facultad de Cornell en 1967. En 1973 y 1974 tuvo la oportunidad de trabajar en la Universidad de Oxford. El tema principal de su investigación, que se refleja en los trabajos descritos en su artículo, es el papel que desempeñan en el comportamiento aquellas neuronas que pueden identificarse una a una.

DAVID EMIL THOMAS ("La formación de imágenes en los espejos") se licenció en matemáticas y física y doctoró en matemáticas por el Instituto de Minería y Tecnología de Nuevo México. Su tesis doctoral se refería a la óptica de espejos con curvaturas arbitrarias. Se ha interesado también por el estudio de la geometría diferencial de superficies, las proyecciones geométricas y la perspectiva cilíndrica del artista holandés M. C. Escher.

Hace...

José M.ª López Piñero

... *cuatrocientos cincuenta años*

Falleció en Valencia Pere Bernat d'Olesa i Rovira, máxima figura española del atomismo renacentista. Durante casi un siglo ha circulado entre los historiadores de la filosofía y de la medicina españolas la noticia de que un tal Pedro Dolese, "médico y caballero valenciano", autor de una obra titulada *Summa philosophiae et medicinae,* fue uno de los más tempranos representantes del atomismo en la España del siglo XVI. El origen de la misma se encuentra en un estudio de Menéndez Pelayo (1879), cuya única información era una breve referencia del atomista Isaac Cardoso en su *Philosophia libera* (1673). Sobre esta base, el historiador santanderino convirtió a "Dolese" en un punto de partida de una tradición atomista genuinamente española, representada después por Gómez Pereira y Francisco Valles y que durante los siglos XVII y XVIII enlazó con los seguidores que en nuestro país tuvo el atomismo de Pierre Gassendi y de Emmanuel Maignan. Esta noticia ha sido reproducida más tarde en numerosas ocasiones, sin que se hubiera localizado la obra en cuestión y sin dato alguno acerca de la personalidad de su autor. La desconexión entre los historiadores y la erudición local impidió que se aprovechara la breve referencia acerca del mismo contenida en la *Biblioteca de Escritores Baleares* (1868) de Joaquín María Bover.

En 1968 nos fue posible localizar un ejemplar del libro, así como reunir algunos datos documentales en archivos mallorquines y valencianos que proporcionaron notable luz acerca de la biografía de "Petrus Dolese", es decir, del médico mallorquín afincado en Valencia Pere Bernat d'Olesa i Rovira.

Hijo de un caballero que fue "conseller" de la ciudad de Palma de Mallorca, estudió artes y medicina en la Universidad de Pisa de 1490 a 1495. Este último año, debido a la llegada del ejército de Carlos VIII de Francia, se trasladó a Montpellier y desde allí, a causa también de la guerra, a la Universidad de Lérida. En 1497, concluidos ya sus estudios con los títulos de doctor en artes y medicina, marchó a Valencia, ciudad en la que permaneció hasta su muerte.

En Valencia alcanzó gran prestigio profesional y científico. A partir de 1497, ocupó el cargo de "examinador de médicos" durante muchos años. Incluso fue nombrado titular de una de las cátedras de medicina de la Universidad, puesto al que tuvo que renunciar en 1514 a causa de sus ocupaciones como clínico. Dictó su testamento en julio de 1530 y murió en los primeros días de noviembre de 1531. De acuerdo con el testimonio de su discípulo Francesc Pujades, treinta días antes de su fallecimiento terminó de redactar su obra, en la que había estado trabajando durante más de dos décadas. El original fue acabado de transcribir por el propio Pujades en 1535 y al año siguiente fue impreso a expensas de Gaspar d'Olesa, primogénito del autor. Bover habla de otra edición anterior (1532) que, caso de existir, no pudo incluir el tercer tratado de la obra.

La *Summa totius philosophiae et medicinae* (1536) es un grueso volumen de más de cuatrocientas páginas, dividido en tres tratados, cada uno con foliación propia. Incluye dos cartas de Gaspar d'Olesa a Fernando de Aragón, Duque de Calabria, personaje bien conocido del Renacimiento valenciano.

El primer tratado está consagrado principalmente a exponer las ideas de Olesa sobre filosofía natural. Rechaza la doctrina aristotélica de la materia prima y las formas sustanciales, oponiéndole una terminante concepción atomista: "elementa sunt ingenerabilia et incorruptivilia et minima divisa; et per spheras disseminata". Critica asimismo agriamente el esquema aristotélico y galénico de las cualidades y afirma que éstas dependen exclusivamente de los "corpuscula" que componen los mixtos. Se interesa, no sólo por los pares "caliditas et frigiditas", "humiditas et siccitas", sino muy especialmente por "gravitas et levitas, luciditas et opacitas et diaphaneitas et colores". Aplica después esta concepción atomista al estudio de los mixtos, tanto minerales, vegetales, como animales, y también a un detenido análisis de las sensaciones. A lo largo de esta exposición se apoya explícitamente en las doctrinas de Anaxágoras y Demócrito.

El tratado segundo tiene diez capítulos, dedicados todos ellos a la aplicación a la medicina de la teoría atomista o corpuscular, por atenerse al término preferentemente utilizado por Olesa. Siguiendo la pauta del *Canon* de Avicena –aunque con una actitud muy crítica frente a sus doctrinas– se ocupa de los conceptos de salud y enfermedad y de los epígrafes habituales de la medicina teórica: "de elementis", "de complexionibus", "de humoribus", "de membris" y "de virtutibus". Ofrece interpretaciones atomistas, no sólo de los problemas fisiológicos generales, sino de las funciones concretas. Afirma, por ejemplo, que el calor cardíaco y el de todo el cuerpo es moderado por el aire que entra por los pulmones, gracias a los "frigida corpuscula qua una cum aëre ingrediunt". En anatomía, dispone de una información libresca basada principalmente en Galeno, al que concede mayor autoridad en la materia que a Avicena. Recurre, sin embargo, al lugar común utilizado por los autores que se resistían a atenderse a la experiencia disectiva: "La anatomía del cadáver es muy diferente de la del vivo". En las adiciones a varios capítulos de este tratado, escritas por Olesa al final de su vida, se incluyen comentarios de carácter clínico y terapéutico.

El tratado tercero, mucho más extenso, es una especie de glosa o desarrollo de los temas expuestos en los dos anteriores. Sus nueve primeros capítulos vuelven a exponer con amplitud la filosofía natural atomista. El décimo se titula "actus intelligendi et sentiendi" y los seis últimos se ocupan de los cinco sentidos exteriores, más un apéndice sobre las causas de los sueños, las profecías y las adivinaciones. El doce y el trece son un detenido estudio de los medicamentos como cuerpos mixtos, con epígrafes independientes consagrados a "De actuatione medicinarum" y "De gradibus medicinarum". Los capítulos catorce y quince exponen la doctrina del autor acerca de la generación espontánea.

La pauta expositiva de Olesa y el lenguaje que utiliza son los típicos del escolasticismo arabizado. Además de

Anaxágoras y Demócrito, Aristóteles, Ptolomeo, Hipócrates y Galeno, cita frecuentemente a autores árabes como Avicena, Averroes y Avenzoar y a cristianos bajomedievales como Gentile da Foligno. Un mundo intelectual y una forma de expresión que serían, sin duda, consideradas con desprecio por la nueva generación de médicos y científicos que dominó en Valencia a partir de 1540: la encabezada por Miguel Jerónimo Ledesma, Pedro Jaime Esteve, Pedro Jimeno y Luis Collado, brillantes seguidores del humanismo, la anatomía vesaliana y del nuevo galenismo "hipocratista". Incluso externamente, la impresión gótica y el formato casi de incunable del libro de Olesa contrastan con las ediciones de las obras de estos autores, de impecables tipos griegos y latinos inspirados en los aldinos. Sin embargo, desde una perspectiva adecuada, la *Summa* tiene una significación histórica igual o superior, como exposición sistemática de la filosofía natural, la biología y la medicina desde los supuestos atomistas. No resulta extraño que fuera redescubierta por Cardoso, seguidor del atomismo de Maignan y Gassendi.


¶ Cũ priuilegio impiali.

¶ Incipit Sũma totius philosophie ⁊ medicine Magnifici domini Petri Dolese equitis: artiũ ⁊ medicine doctoris celeberrimi: que in tres principales tractatus diuiditur: ⁊ incipit primus/qui primus tractatus Sũme medicine appellatur. In quo multa pertractantur que ad philosophiam expectant: ⁊ multa dimittũtur medicinalia i aliis duobus tractatibus sequentibus pertractãda. Anno salutis christiãe. M.D.xxxvi.

Una de las portadas de la **Summa totius philosophiae et medicinae** *(1536) de Pere d'Olesa, principal exposición sistemática del atomismo en la España del siglo* XVI

La fábula

Desde el año 2500 a. C. hasta nuestros días, la fábula se ha mantenido como un género popular y crítico. Podemos reconstruir ya su historia, dentro de una tradición fundamentalmente unitaria

Francisco Rodríguez Adrados

Nuestro lenguaje, todavía hoy, está lleno de alusiones a fábulas y de proverbios animales que, en definitiva, provienen de la fábula greco-latina. Hablamos de "vestirse con piel de cordero" o "con plumas ajenas", de que "aunque la mona se vista de seda, mona se queda", de "la parte del león" o "la gallina de los huevos de oro". Y seguimos atribuyendo a animales como la serpiente, la zorra, el mono, el lobo, etcétera, las cualidades que la fábula desde tiempos antiguos les atribuye.

Todo esto testimonia la vitalidad y popularidad de un género que permanece esencialmente idéntico a sí mismo desde la antigua Sumeria, en la segunda mitad del tercer milenio antes de Cristo: en Mesopotamia, en la India, en Grecia, en Roma, en nuestra Edad Media, en la Edad Moderna. Hoy podemos reconstruir aproximadamente la historia de este género gracias a nuevos descubrimientos e investigaciones. Es de esta historia de la que queremos presentar aquí un panorama, en términos escuetos.

Quizás el modo más directo de hacer patente esa continuidad esencial de la fábula dentro del mundo cultural a que nos hemos referido, sea recoger unas pocas, poquísimas fábulas de distintas edades y lugares dentro del mismo. Luego pueden obtenerse, como conclusiones, unos pocos rasgos comunes, que se amplían con otros datos más. Esto puede servir de prólogo al estudio que sigue sobre el detalle de la historia de la fábula: de cómo ha habido una difusión de edad en edad y de cultura en cultura, con lo que se explica la uniformidad del género, de la cual partimos.



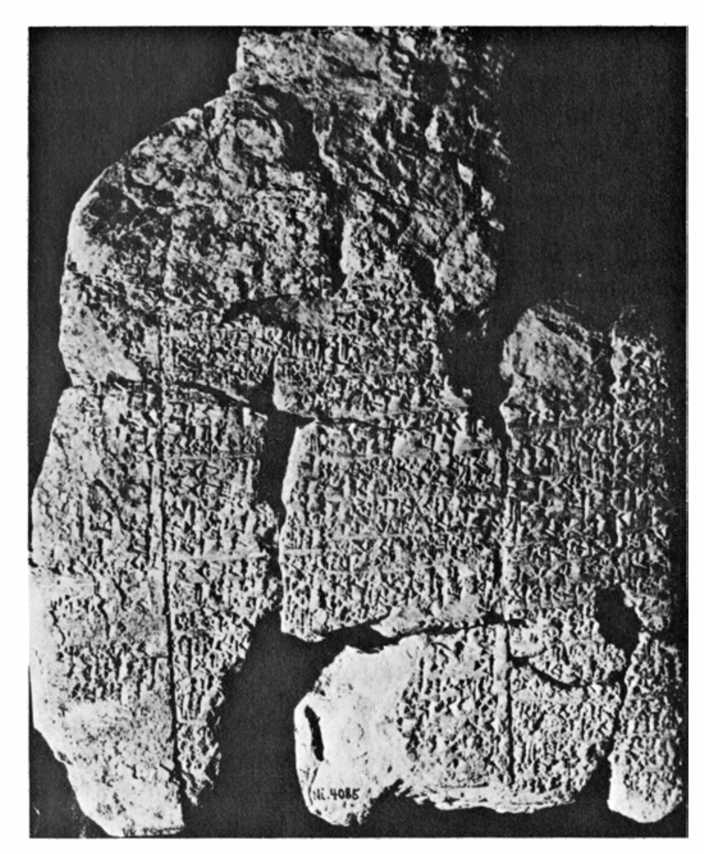
ESOPO Y LA ZORRA, representados en un vaso ático de figuras rojas del Museo Vaticano, fechado en el siglo v a. C. La fábula griega la conocemos desde Hesiodo ("el ruiseñor y el halcón"), en el siglo VIII.

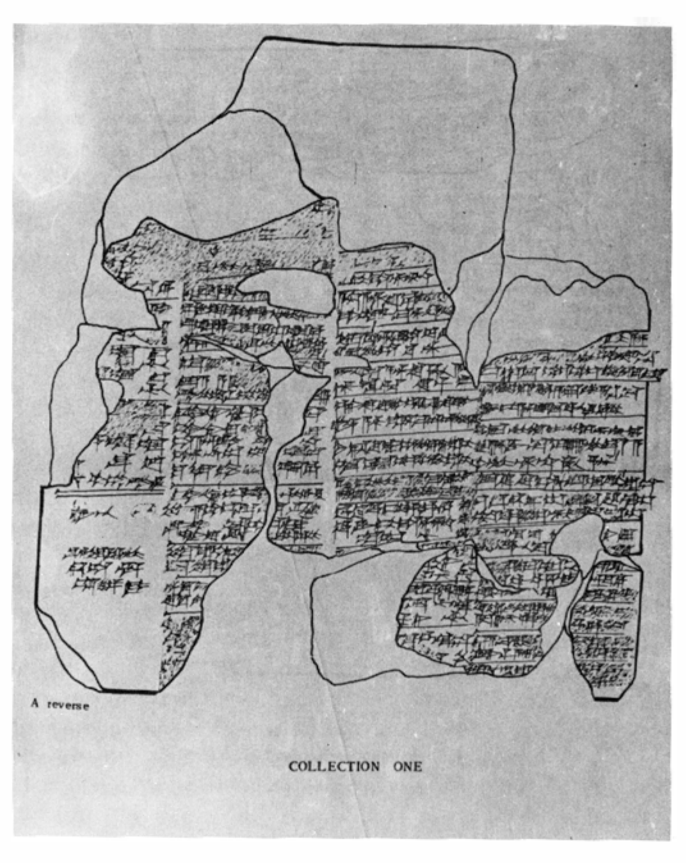
En Mesopotamia la fábula remonta al período de la segunda mitad del tercer milenio antes de Cristo (el león presidiendo como rey a los animales se encuentra en una lira de la primera dinastía de Ur, hacia el 2500 a. C.), pero lo que ha llegado a nosotros son textos del período neo-sumerio (hacia el 2000-1900) que fueron copiados en el babilonio antiguo (hacia 1700-1600 a. C.). Hay fábulas en lengua acadia, asiria y neo-babilonia: desde la fecha mencionada hasta el siglo VI a. C. He aquí tres fábulas mesopotámicas: la primera sumeria, las dos últimas asirias.

El león y la cabra. El león había apresado a una indefensa cabra. "Dejame ir –dijo la cabra– y te daré una oveja amiga mía". "Si quieres que te deje ir –dijo el león– dime tu nombre". La cabra le respondió: "¿No conoces mi nombre? 'Tu eres inteligente'". Cuando el león llegó al redil, dijo: "Ahora que he llegado al redil, voy a soltarte". Y la cabra le respondió, cuando estuvo al otro lado de la valla: "¡Me has soltado! ¿Eras tan inteligente? En vez de darte la oveja, voy a huir".

El mosquito y el elefante. Un mosquito, después que se hubo posado sobre un elefante, dijo: "Hermano, ¿te peso mucho? Me bajaré en la charca". El elefante replicó: "Me da igual que sigas. ¿Qué más me da llevarte? Ni tampoco me importa que te marches".

Las golondrinas y el gato. Las pequeñas golondrinas se cayeron del nido; y





FRAGMENTO DE TABLILLA que recoge la primera colección sumeria de proverbios y fábulas. El dibujo reproduce el texto anterior. Las fábulas mesopotámicas se remontan a la segunda mitad del tercer milenio antes de Cristo, aunque los textos que nos han llegado corresponden al período neo-sumerio.

un gato las cogió y les dijo: "Si no hubiera sido por mí, una gran desgracia os habría sucedido". Ellas respondieron y le dijeron: "¿Es por eso por lo que nos pones en tu boca?".

La fábula griega la conocemos desde "El ruiseñor y el halcón", en Hesíodo (siglo VIII a. C.); luego en diversas fábulas que sirven de ejemplo en poetas como Arquíloco (s. VII), Semónides (s. VI), cómicos como Aristófanes (s. V); otras se atribuyen a Sócrates y a los socráticos en general. Luego, a partir del año 300 a. C. se recogen en colecciones, la más antigua la de Demetrio de Falero; y se crean, por supuesto, nuevas fábulas. Muchísimas fábulas griegas aparecen, más o menos modificadas, en autores latinos: Fedro (s. I d. C.) y Aviano (s. V d. C.) en la antigüedad, otros en la Edad Media. He aquí unas pocas fábulas de esta tradición, en versión resumida: la primera en Arquíloco (luego está en las colecciones), la segunda en la más antigua colección griega que nos ha llegado (la Augustana, cuyo texto conservado es del s. V d. C., pero reelabora otros anteriores), la tercera en Fedro.

La zorra y el mono. El mono y el camello pretendían reinar sobre los animales. El mono censuró el temperamento pacífico del camello y triunfó de él en el baile; el camello fue expulsado a palos y el mono proclamado rey, vistiéndose sandalias doradas, una piel de león y un manto de púrpura. La zorra, envidiosa, se le acercó y le señaló un trozo de carne puesto en un cepo, diciéndole que le correspondía como rey. Así le engañó y el mono fue aprisionado por el cepo, quedando en una posición indecorosa. La zorra le dijo, en son de burla: "¿Pretendías ser rey teniendo, oh mono, un trasero como ese?".

El cuervo y la zorra. El cuervo arrebató un trozo de carne y se posó en un árbol. La zorra se le acercó y elogió su belleza; dijo que debería ser rey de las aves y que lo sería sin duda si supiera cantar. El cuervo intentó cantar, con lo que dejó caer la carne mientras graznaba. La zorra la cogió y dijo: "Oh cuervo, si tuvieras sensatez nada te faltaría para ser rey de los animales".

El lobo y el cordero. El lobo y el cordero, sedientos, llegaron al mismo arroyo. El lobo estaba aguas arriba del cordero. Incitado por el hambre, buscó causa de pelea. "¿Por qué –dijo– has ensuciado mi agua?". Y el cordero, asustado: "¿Cómo puedo, lobo, hacer eso de que te quejas? El agua viene de ti a mi boca". Pero el lobo, derrotado por la verdad, replicó: "Hace seis meses hablaste mal de mí". Respondió el cordero: "Yo no había nacido todavía". "Pues entonces fue tu padre, por Hércules, el que habló mal de mí", contestó el lobo. Y cogiéndole, le despedazó con muerte injusta.

Si pasamos a la fábula india, ésta se encuentra ya a manera de ejemplo en distintas obras literarias (como, por ejemplo, el poema épico *Mahābhārata*) ya en colecciones de las que la más difundida es el *Pañchatantra,* muy reelaborado en distintas versiones desde el siglo V al XIII d. C. y que amplía considerablemente una colección anterior, el llamado *Tantrākhyāyika,* cuya redacción pensamos debe situarse en el siglo II a. C. He aquí, ahora, tres fábulas indias en versión resumida.

La tortuga y los cisnes. A punto de secarse un lago, dos cisnes que vivían en él se disponían a marchar a otro más grande, pero antes se despidieron de su amiga la tortuga. Esta se quejó de que la dejaran expuesta a una muerte segura y les pidió que la llevaran con ellos. Los cisnes accedieron: sujetaron por ambos lados una vara de la que debía cogerse la tortuga con los dientes; mientras fueran por el aire, no debía hablar una sola sílaba. Así hizo hasta que los tres pasaron por encima de la ciudad, cuyos habitantes dijeron "¿Qué es esa cosa del tamaño de una rueda de carro que es llevada a través del aire?". La tortuga al oírlo soltó la

vara y dijo: "Soy una tortuga. Esos hombres dicen tonterías". Pero al hablar se cayó y fue cogida por los hombres, que la despedazaron y se la comieron.

El asno con piel de pantera. Un asno yacía desfallecido de llevar tan pesadas cargas. El amo imaginó ponerle una piel de pantera para que pudiera reponerse comiendo en abundancia de los campos de los vecinos, que se asustarían. Así hizo y, efectivamente, un labrador que vio lo que creía era una pantera, emprendió la fuga arrastrándose: llevaba en sus manos un arco y se cubría con un manto. Pero al verlo huir de esa manera el asno le confundió con una burra y le persiguió, con lo que el labrador comenzó a arrastrarse más deprisa. El asno entonces pensó que la burra huía por creer que se las había con una pantera: para atraerla, tiró la piel de pantera y se puso a rebuznar. El labrador, entonces, se puso en pie, se dio la vuelta y mató al asno con su arco.

El gato asceta. La perdiz y la liebre, que tenían una disputa porque la segunda había ocupado la casa abandonada por la primera, buscaron un juez que decidiera su caso. Acudieron a un gato asceta que permanecía en pie a orillas del río, haciendo penitencia. El asceta elevó las manos al cielo y comenzó a orar y a pronunciar sentencias morales, ganando así la confianza de los pleiteantes. Estos se acercaron y fueron devorados.

Vamos a limitarnos a narrar una fábula medieval, latina, independiente de la tradición antigua, para que se vea que, sustancialmente, hay coincidencia.

El lobo aprendiendo a leer. Un sacerdote enseñó a un lobo las letras. Dijo "A" y el lobo igual; y lo mismo con las otras letras. "Unelas ahora y silabea", dijo el sacerdote. "Aún no sé silabear", contestó el lobo. Y el sacerdote: "Hazlo como mejor te parezca". Y dijo el lobo: "Lo que mejor me parece es lo que suena *agnus* (cordero)". Y el sacerdote dijo: "Lo que está en el corazón, está también en la boca".

No sería difícil, a partir de este breve material (expuesto, por otra parte, en forma sumaria), deducir algunos rasgos comunes. Nos hallamos ante relatos ficticios en que hablan los animales; otras veces se introducen también plantas y objetos diversos, pudiendo intervenir también hombres o dioses. Hay, por lo demás, relatos en que sólo entran hombres o dioses, es decir, los géneros de la anécdota y el mito apenas se distinguen. Se refieren a acciones que supuestamente sucedieron una vez, en tiempos pasados, y que se dan como ejemplos de valor general; el que las fábulas se recojan en colecciones y lleven, ocasionalmente, moralejas, es un proceso secundario.



LA FABULA DEL CUERVO Y LA ZORRA nos ha llegado a través de la colección Augustana, la más antigua de las griegas, cuyo texto conservado es del siglo v d. C. La ilustración es de Grandville (s. XIX).

Los personajes de la fábula actúan como símbolos de distintos caracteres humanos; la acción que sucede entre ellos es paradigmática de lo que sucederá en cualquier tiempo. Porque la naturaleza es constante: el disfraz o la condición social o la edad, etc. no cambian al animal, ni tampoco, se entiende, a los hombres. La concepción de la vida humana que subyace a la fábula es realista y crítica. El poderoso se impone, sean cualesquiera las razones del débil, y no hay en él caballerosidad, cuanto más hipocresía. Pero el débil puede ser superior en ingenio y triunfar con el engaño o la astucia. Y hay crítica y burla de la vanidad, la tontería, la codicia.

Nos hallamos ante un género popular y satírico, nada idealizante. Los rasgos cómicos y las salidas chistosas están en él a la orden del día, pero esto no oculta, sino al contrario, la dureza de la vida. Esta es concebida como un constante enfrentamiento: la mayoría de las fábulas son esto, trátese de enfrentamientos de acciones o verbales, sea el resultado un triunfo efectivo o una salida chistosa que da, en cierto modo, la "solución".

Aunque a veces la fábula es de "situación", da el comentario del animal víctima de un poder superior o de su propia necedad, y otras veces es "etiológica", explica el porqué de ciertos hechos a partir de un suceso pasado, no por ello dejan de estar presentes los mismos rasgos fundamentales: visión cruel de la vida humana, constancia de la naturaleza, sátira. He aquí una fábula de "situación", del poeta griego Semónides, y una fábula "etiológica" (origen del moño de la abubilla) de Aristófanes: *a*) *El pescador y el pulpo:* "Viendo un pescador un pulpo en el agua en un día de invierno, se dijo a sí mismo: 'Si no me echo a nadar, voy a pasar hambre'". *b*) *La alondra:* "Esopo dice que la alondra nació antes que las demás aves, antes que la tierra, y que después su padre enfermó y murió; y no había tierra y el cadáver estuvo expuesto cinco días; y ella, apurada ante la dificultad, enterró a su padre en su propia cabeza".

Por lo demás puede haber múltiples variantes: presencia o ausencia del diálogo, acción agonal simple o doble, aparición de un personaje ajeno a la fábula y que saca su conclusión, etc. etc. Y la fábula, en las distintas literaturas, puede aparecer ya como un simple ejemplo (mezclado con otros de carácter mítico, anecdótico, novelístico

y con símiles, proverbios, etc.), ya de otras maneras varias. Señalemos las principales.

1. *Ejemplos contados "en segundo grado"*. Es cuando el que va a contar una fábula lo hace poniéndola en boca de alguien en una situación parecida a la suya. El caso más conocido es el de poner las fábulas en boca de Esopo, personaje mítico que argumentaba con fábulas en diversos episodios de su vida y, finalmente, cuando fue condenado a muerte injustamente por los delfios.

Así, si Semónides, en el s. VI, contó de por sí la fábula de "El águila y el escarabajo" (el sucio escarabajo puso su bola en el regazo de Zeus, donde había depositado sus huevos el águila: el dios lo sacudió, con lo que tiró al suelo los huevos y el escarabajo quedó vengado), cuya intención es hacer ver que no hay enemigo pequeño, Aristófanes nos presenta en sus *Avispas* al personaje Filocleón que, llamado a juicio por sus desafueros, relata lo que dijo Esopo (la fábula en cuestión) en el episodio aludido, según él semejante a su propia situación. A partir de aquí la atribución de las fábulas a Esopo ha sido frecuentísima.

2. *Colecciones*. Un tipo de colección coloca las fábulas (y a veces, proverbios, anécdotas, etc.) simplemente unas detrás de otras; así en diversas colecciones sumerias y babilonias y así en Grecia a partir de Demetrio de Falero. Esta tradición griega es la más generalmente seguida en adelante; dado que las fábulas van en ella fuera de cualquier contexto, es frecuente que se provean de moralejas explicativas, iniciales (promitios) o finales (epimitios).

Pero hay otro tipo de colección, aquel en que aparece una acción que hace de marco de las fábulas. Es el tipo indio bien conocido. En el *Pañchatantra,* un rey encarga a un brahmán la educación de sus rebeldes hijos y éste la realiza por medio de fábulas. En realidad se trata de cinco fábulas, dentro de cada una de las cuales los personajes argumentan a su vez con otras fábulas. Este tipo fue muy imitado en la Europa medieval (Pedro Alfonso, el *Conde Lucanor* de Don Juan Manuel, etc.).

3. *Vidas*. En la *Vida de Ahikar,* asiria, este secretario del rey Senaquerib da consejos a su hijo adoptivo Nadan por medio de proverbios y fábulas y luego, cuando se porta villanamente, le reprende de igual modo. En realidad, se trata de un precedente de las colecciones de tipo indio. En Grecia se atribuyen fábulas a Esopo en diferentes circunstancias de su vida; y conservamos una *Vida de Esopo* cuyo original remonta a la edad helenística y que está llena de fábulas, proverbios, símiles, etc. con que Esopo argumenta frente a sus contrincantes. Recordemos en nuestra Edad Media obras como *El Libro del Buen Amor* del Arcipreste de Hita, en que Trotaconventos y Doña Garofa argumentan también con ayuda de fábulas sobre el tema de si la segunda debe o no aceptar las proposiciones del Arcipreste.

PAGINA INICIAL de la edición de *Tantrākhyāyika*, de Herter (Leipzig, 1914). Esta colección de fábulas indias, compilada en el siglo II a. C., constituye el precedente de la compilación del *Pañchatantra*.

4. *Epopeya animal*. En Grecia hay una epopeya burlesca, la *Batracomiomaquia* o lucha de las ranas y los ratones que, en realidad, es una fábula ampliada. En la Edad Media, de un modo paralelo, diversas fábulas de la tradición antigua en torno al zorro han dado epopeyas animales latinas como la *Ecbasis captivi* o fuga del ternero cautivo del lobo (s. X) y el *Ysengrimus* (s. XII), sobre enfrentamientos del lobo y la zorra; el *Roman de Renart* francés es una derivación de estas obras.

Los distintos tipos de fábulas, las diversas maneras de presentarlas, se repiten a lo largo de una historia que va del tercer milenio a. C. a nuestros días. Naturalmente, los tipos de sociedad varían y la crítica se dirigirá ya a los poderosos mesopotámicos, griegos o indios, ya a los filósofos griegos, los brahmanes indios o los nobles y el clero medieval. Hay también la moralización de la fábula, su adaptación a ciertos modelos, su censura. Pero sigue siendo cierta su fundamental identidad. Quizá la mejor manera de mostrar esto sea trascribir dos fábulas nuevas, que la revista *Newsweek* recogió de un discurso del banquero americano Bert Lance, asesor económico del ex-presidente Carter caído en desgracia por alegaciones sobre irregularidades económicas.

Es el eterno tema del inocente que sufre abuso y de la "paja en el ojo ajeno".

El lobo y el cordero. Un paseante veía todos los días un lobo y un cordero que estaban pacíficamente el uno al lado del otro en el jardín de una casa. Intrigado, preguntó al amo como conseguía ese milagro. "Pongo un cordero nuevo cada día", respondió.

La señora, el pollero y el pollo. La señora palpó uno a uno los pollos de una pollería e iba ya a marcharse porque no encontraba ninguno de su gusto cuando el pollero le dijo: "Señora, ¿está segura de que Vd. misma saldría airosa si fuera sometida a una prueba tan severa?".

Que la fábula griega tiene relación con la fábula oriental, no era dudado en la antigüedad, que atribuyó a Esopo un origen frigio o tracio. El fabulista Babrio, a comienzos del s. II d. C., llegó a más diciendo que la fábula tiene su origen en Asiria: esto es, en Mesopotamia.

Sin embargo, por causa de nuestro desconocimiento de la fábula mesopotámica hasta hace pocos años, durante mucho tiempo todas las discusiones sobre el origen de la fábula se limitaban al tema de si la fábula griega procedía de la india o viceversa. Benfey, el editor del *Pañchatantra* en 1859 y verdadero fundador de la literatura comparada, se inclinó por la tesis del origen griego, que es la que luego fue más generalmente seguida; un indianista como Edgerton sostuvo, en cambio, el origen indio. Hausrath, uno de los primeros estudiosos alemanes de la fábula, creyó posible demostrar el origen griego de fábulas indias allí cuando existe una versión de las mismas en fuentes griegas arcaicas.

Pero no pasaron de unos poquísimos casos aquellos en que Hausrath hallaba una relación estrecha entre fábula griega y fábula india. Uno de los más notables es el de "El león, la zorra y el ciervo": cuando la zorra llevó al ciervo a la caverna del león y éste le echó la zarpa, hiriéndole en la oreja, el ciervo huyó asustado, pero la astuta zorra logró que volviera, diciéndole que se trataba de una caricia del león. Naturalmente, la fiera devoró al ciervo, pero cuando iba a comerse el corazón no lo encontró: y la zorra, que lo había devorado, dijo: "¿Cómo podría tener corazón un animal que entró dos veces en la caverna del león?". La versión india está muy próxima, sustituyendo al ciervo por el asno y la zorra por el chacal. Con todo, aparte de aducir ejemplos muy escasos, Hausrath no contaba con una hipótesis al menos posible: que ambas versiones derivaron de Mesopotamia.

Hoy que conocemos bastantes fábulas mesopotámicas, sabemos que esta hipótesis es, en muchos casos al menos, acertada. No quiere esto decir que no hubiera en Grecia una tradición fabulística antigua. La había: la fábula es una variante popular, satírica, del mito y se contaba en fiestas agrarias de tipo popular. Ello se demuestra, entre otras cosas, porque en Grecia aparece ligada desde el s. VII al género yámbico, caracterizado por el uso del pie yámbico (grupo de una sílaba breve y una larga) y por el carácter festivo, satírico y libre. Arquíloco y Semónides, entre otros, son escritores de yambos que incluyen fábulas en sus poemas.

Pero no hay duda de que en el momento en que en Grecia las fábulas que se contaban en las fiestas populares y los banquetes comenzaron a tomar forma literaria, a ser escritas por los poetas, esto no sucedió sin un fuerte influjo oriental. Bastarán unos pocos ejemplos.

La fábula del águila y la zorra, en Arquíloco, cuenta cómo estos animales hicieron amistad, pero el águila violó el juramento prestado comiéndose las crías de la zorra. El águila se reía, en un picacho, de los lamentos y amenazas de la zorra: pero ésta acudió al dios Zeus, defensor del juramento. El águila sacrílega cogió carne de un sacrificio, llevándose con ella unos carbones encendidos. El resultado fue que se prendió su nido y perecieron sus polluelos: así fue vengada la zorra por Zeus. Pues bien, desde hace algunos años sabemos que esta fábula deriva de una fábula-ejemplo que aparece en la epopeya acadia *Etana:* el tema es muy aproximado, aunque en vez de la zorra aparece la serpiente y es el dios Samas el que hace el papel de Zeus. El águila es el servidor del dios (como lo es también de Zeus) y es castigada por él por su traición a la serpiente.

No es este el único ejemplo. Hay paralelos mesopotámicos muy exactos a fábulas griegas como "Los árboles y la caña", "La zorra y las uvas", "La víbora y la lima", "El ratón agradecido", "El caminante y el plátano", "Zeus y el camello" (que pide cuernos y es castigado), "El toro y el mosquito" (es la fábula asiria antes mencionada de "El elefante y el mosquito") y otras muchas más. Puede haber pequeñas diferencias: en vez de la zorra y las uvas en Mesopotamia se trata del perro y los dátiles, el toro es sustituido por el elefante, el camello por el asno. Pero la sustancia es la misma.

Y no se trata sólo de esto. La caracterización de los animales (tema del reinado del león, por ejemplo), los tipos de acción (por ejemplo, la disputa entre dos plantas), la estructura misma de las fábulas, los temas de la naturaleza, de la imposición del fuerte, del



LOS DELFIOS DESCUBREN LA COPA escondida en el equipaje de Esopo, sobre quien recaerá injusta condena. Se trata de un grabado que ilustraba la *Vida de Esopo* en el incunable de Zaragoza, de 1489.

ingenio triunfante, son los mismos. De otra parte, es claro que el personaje griego Esopo, narrador de fábulas, ha sido conformado sobre el modelo de Ahikar, el secretario del rey asirio Senaquerib, narrador de fábulas y proverbios, solucionador de enigmas, víctima de calumnias que le llevan a una muerte que, en este caso, es sólo aparente.

Llegados a este punto, surge la hipótesis de que la fábula india puede, a su vez, haber recibido influjo de la fábula mesopotámica. Esto no quiere decir que no haya raíces indígenas en la fábula india: en realidad, se trata de un género que surge independientemente en todas las culturas. Ni niega tampoco que en la India se hayan producido adaptaciones: por ejemplo, la introducción de animales indígenas como el tigre o la presencia de ambientes o de tipos humanos propiamente indios.

Pero ha debido de haber, además, un influjo mesopotámico: influjo que se establece claramente cuando hay coincidencias entre fábulas mesopotámicas y fábulas indias concretas; y que se establece indirectamente cuando hay coincidencias entre fábulas indias y fábulas griegas arcaicas o clásicas. La relación entre las culturas de la India y la de Mesopotamia ha existido en una serie de períodos. Y esto lo mismo en fecha antigua, cuando imperaban en Babilonia los Kasitas, una dinastía indo-irania, desde el s. XVIII, que en otra más reciente, cuando en tiempos de Darío, en los siglos VI y V, el valle del Indo era una satrapía persa y Mesopotamia estaba, igualmente, bajo el dominio persa.

Pongamos ahora algunos ejemplos que hagan verosímil la hipótesis sentada. El tema de la fábula acadia "El águila y la serpiente" se encuentra prácticamente en dos fábulas indias del *Tantrākhyāyika* en que intervienen ya dos cornejas y la serpiente, ya el cangrejo y la serpiente. Las fábulas indias respetan la serpiente (sustituida por la zorra en la versión griega) y sustituyen, en cambio, el águila. También otras fábulas presentes en Mesopotamia y Grecia se encuentran igualmente, con más o menos variantes, en la India. Este es el caso de "El águila y la tortuga" (tema de la naturaleza, la tortuga sufre las consecuencias de querer volar), de "La encina y la caña", de "El elefante y el mosquito" (aquí se trata del león, que desdeña matar al mosquito).

En otras ocasiones, como digo, no hay versión mesopotámica conocida: pero las coincidencias entre una fábula griega arcaica y una india hace verosímil la hipótesis del origen mesopotámico de ambas. Así en el caso de "El león, la zorra y el ciervo", ya tratado, y en otros varios más. Otras veces son los temas, los motivos de una fábula los que coinciden. Por ejemplo, la fábula de Arquíloco del león y la zorra (la zorra no quiere entrar en la cueva del león porque ve huellas de animales que entran, pero no de animales que salen) aparece en la India descompuesta en dos. En una fábula, el chacal hace astutamente que el león traicione su presencia en la cueva en que se esconde; en otra, un mono se escapa de caer en poder de un demonio mediante una observación sobre las huellas de diversos animales idéntica a la de la zorra de Arquíloco.

Pero no es sólo esto. En la fábula india aparecen temas que son familia-



ESOPO Y ALGUNOS DE SUS ANIMALES Y PERSONAJES. **Ilustración perteneciente al incunable de Zaragoza referido antes. Era frecuente poner en boca de este mítico personaje muchas fábulas.**



LA ZORRA Y EL GALLO. La fábula se inscribe en lo que el autor denomina epopeya animal, de gran difusión en la Edad Media europea. La ilustración pertenece a un códice de la catedral de Gerona (s. x).

res en la griega: por ejemplo, el del águila agradecida, el del animal cuya naturaleza no cambia con el disfraz (en Grecia, el asno se disfraza de león, en la India el chacal de pantera y hay otras variantes más), el del ratón agradecido, el del animal transformado en mujer (la comadreja en Grecia, la rana y el ratón en la India) y que descubre al final sus antiguos instintos, etc.

Hay que añadir el tema del reinado del león, el del chacal servidor de aquel (como la zorra en Grecia) y que se las arregla astutamente para salir de dificultades; y otros más. Se trata de rasgos mesopotámicos, que se ven, por ejemplo, en la "disputa" acadia del león, el perro, el lobo y la zorra.

Todavía más. Ya hemos dicho arriba que es característica de las colecciones indias de fábulas el que éstas sean expuestas dentro de un marco. Se instruye o aconseja por medio de fábulas o dos personajes discuten con ayuda de fábulas. Esto ocurre no sólo en el *Pañchatantra,* sino en modelos de colecciones medievales como el *Sendebar, Las Mil y Una Noches,* el *Conde Lucanor,* etc. Pues bien, ya hemos dicho que en el *Ahikar* asirio hay un precedente de este proceder. No sólo aconseja o reprende Ahikar a su hijo adoptivo Nadan con ayuda de fábulas, sino que, al actuar como consejero del rey de Asiria y resolver los problemas propuestos por el rey de Egipto, representa al tipo humano del filósofo consejero de reyes de las colecciones indias y medievales. No parece dudoso que el *Ahikar* y otras obras mesopotámicas semejantes sirvieran de modelo a los redactores de las colecciones indias.

Como hemos dicho, la fábula griega representa una mezcla de tradiciones indígenas y orientales. En época arcaica y clásica se narra la fábula ya como un ejemplo en "primer grado", ya en "segundo grado" como algo que dijo un determinado personaje (Esopo en general, pero no sólo él: un sibarita, Demóstenes, etc.) en una circunstancia semejante a la que lleva a narrar la fábula.

Son sobre todo los poetas que escriben en yambos, género poético eminentemente popular y abierto a la crítica y la sátira, los que más usan de la fábula: Arquíloco, Semónides, los cómicos como Aristófanes. Ya hemos dicho que esta poesía nació en un ambiente religioso y festivo, de tipo "marginal" en cierto modo. Y es fácil poner la fábula animal y vegetal en relación con estas fiestas: danzas animales, determinadas creencias en relación con los animales, etcétera. Pues bien, cuando la fábula fue adoptada por los socráticos, hay que poner el hecho en relación con el carácter inconformista de este movimiento, su crítica de las formas convencionales de comportamiento, de los valores basados en la riqueza y el poder.

Esto es importante, porque nos explica que fuera un discípulo de Aristóteles, Demetrio de Falero, el creador de la primera colección de fábulas hacia el año 300 a. C. En realidad, Demetrio no hizo otra cosa que poner en la prosa de su época y redactar de nuevo las fábulas que encontró en la literatura arcaica y clásica. Ciertamente, olvidó algunas que conocemos por las fuentes literarias clásicas y que faltan en las colecciones: sin duda por causa de este olvido de Demetrio. Otras veces, fábulas de Demetrio fueron dejadas de lado por las colecciones posteriores (pero las conservó, por ejemplo, la *Vida de Esopo*). También





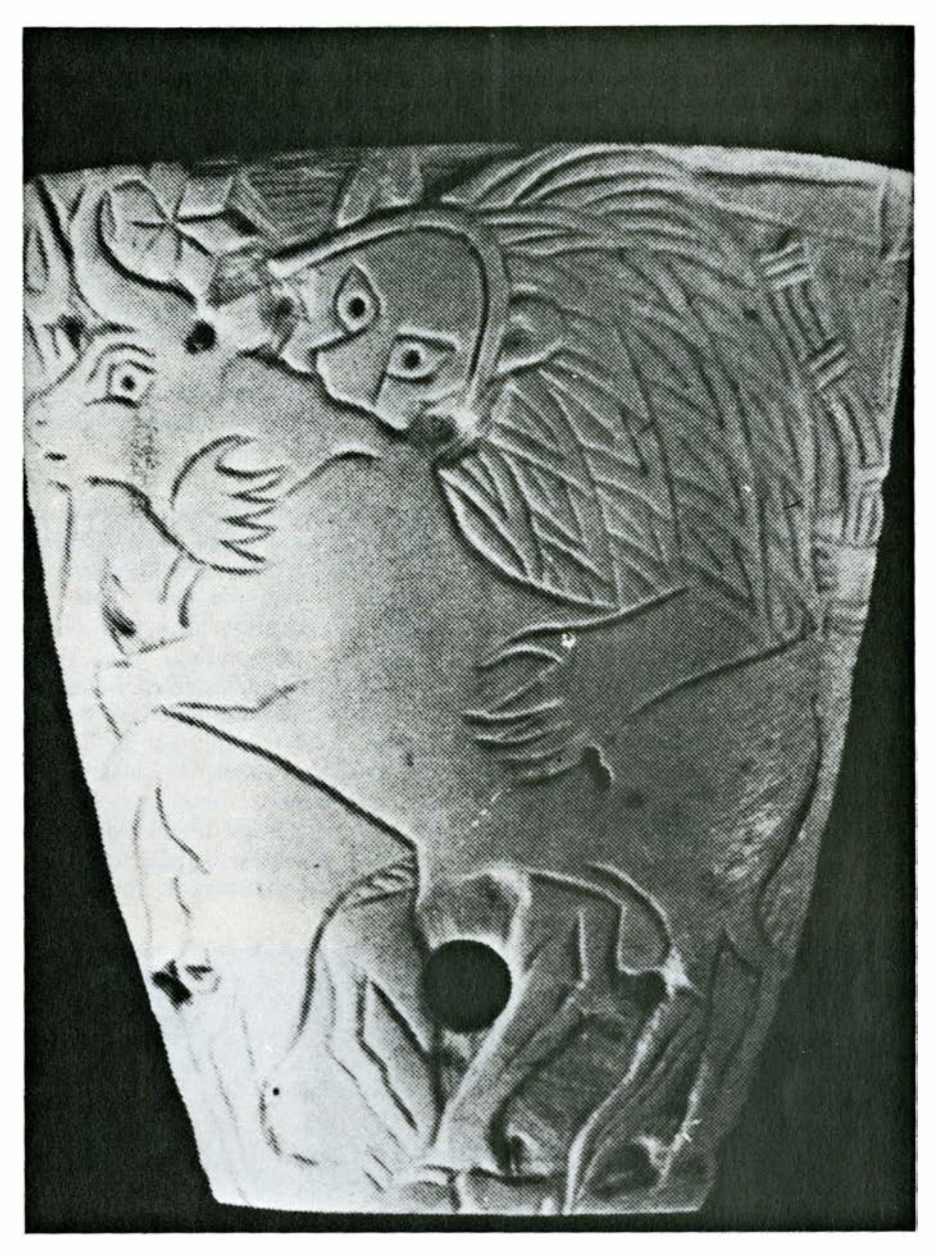
EL AGUILA Y LA ZORRA hicieron amistad; pero el juramento prestado, violado por aquélla, fue vengado taliónicamente por la zorra. Esta fábula y la de la zorra y las uvas tiene paralelos mesopotámicos. (Ilustración de la izquierda, del incunable de Zaragoza. A la derecha, impronta de una gema de Trikka.)

sucede que fábulas que sin duda estuvieron en la colección de Demetrio y de ahí pasaron a las posteriores, son símiles, proverbios, etc., de época anterior que fueron "fabulizados".

La actividad de Demetrio de Falero como autor de una colección de fábulas hay que ponerla en relación con la moda helenística de los "géneros antológicos". Mitos, máximas, epigramas, opiniones de los antiguos filósofos, inscripciones de diversos caracteres, etcétera, se recogían ahora, ya en la redacción original, ya en otra nueva. Aristóteles había favorecido o hecho él mismo colecciones de materiales (sobre victorias musicales o atléticas, datos de zoología o botánica, constituciones de las diversas ciudades, etcétera) que servían luego de base a sus estudios hechos con un método inductivo. Demetrio seguía en la misma línea con un género, la fábula, muy dentro de la tradición socrática. Escribió, igualmente, una colección de *khreiai* o anécdotas que contenían una frase brillante.

La colección de Demetrio no ha llegado a nosotros: sólo la conocemos porque la menciona Diógenes Laercio, el autor de las *Vidas de los Filósofos*. Podemos, sin embargo, reconstruirla hasta cierto punto –es lo que personalmente he intentado en mi *Historia de la fábula greco-latina*–. Esta reconstrucción se basa, primero, en que sólo a través de Demetrio han podido llegar a las colecciones posteriores y a la *Vida de Esopo* una serie de fábulas de la literatura anterior. Y, segundo, en que en esas colecciones posteriores pueden aislarse, por contenido y composición, una serie de fábulas que son a todas luces posteriores a Demetrio.

A veces se puede ir más lejos. Cuando una fábula nos es conocida tanto por la literatura arcaica y clásica como por las colecciones tardías y en éstas presenta rasgos que la diferencian de las versiones antiguas, esos rasgos proceden de una refundición: la de Demetrio. Así, por ejemplo, en "El águila y la zorra". Es claro, en suma, que Demetrio elimina el verso antiguo, sustituyéndolo por la prosa; que elimina también ciertas estructuras complejas, tendiendo a enfrentamientos, *agones,* de un solo acto; que no presenta promitios ni epimitios.



LEON ATACANDO a un toro. La caracterización de los animales (tema del reinado del león, por ejemplo), los tipos de acción (la disputa entre dos plantas) y la propia estructura del relato son parejos en Mesopotamia y en Grecia. La ilustración reproduce una concha grabada procedente de Tello, en Sumeria.

Ahora bien, entre la colección de Demetrio y las colecciones que han llegado a nosotros hay un intervalo de unos 300 años, en el caso más favorable. Es la historia de la fábula griega durante esos años –la época helenística y la romana republicana– la que he intentado reconstruir en el libro arriba citado. El método ha sido la comparación entre las diversas colecciones, que permite reconstruir en cierta medida los estadios perdidos, y el estudio interno de las mismas, sobre todo de la más antigua de las colecciones griegas en prosa, la Augustana.

La colección Augustana contiene más de 250 fábulas, frente a las 100 aproximadamente que debía de comprender la de Demetrio; muchas de ellas son a todas luces postclásicas. Pues bien, la gran sorpresa que ofrece esa colección de fábulas griegas en prosa, que en definitiva es el resultado de una serie de reelaboraciones sucesivas y ampliaciones también sucesivas de la colección de Demetrio, es que, ante un estudio detenido, se revela como prosa griega llena de restos de verso. Concretamente, de versos yámbicos: el llamado trímetro yámbico (seis yambos sucesivos) y el coliambo (idéntico, pero con la sílaba penúltima larga, no breve). En definitiva: en algún momento la colección de Demetrio ha sido versificada en los ritmos aludidos y ha recibido otras fábulas más en los mismos. Y en algún momento esa colección de fábulas en verso ha sido prosificada. Y se han añadido nuevas fábulas totalmente en prosa, sin huella de haber tenido nunca verso. Y se ha provisto a las fábulas, a todas ellas, de epitimios prosaicos, sin huella de verso tampoco.

Este es el descubrimiento cuyos primeros resultados publiqué en la revista "Emerita" en 1969 y 1970 y que he explicado más detenidamente en el libro citado más arriba. En él he hecho ver que el análisis métrico de detalle de

los versos reconstruidos permite ver que se trata de la variante del trímetro yámbico y el coliambo usada en el siglo III a. C., lo que data la versificación. Y he hecho ver igualmente que los restos del metro se conservan sobre todo en "fórmulas" o frases hechas, algunas referentes a temas cínicos.

Es a los cínicos, cultivadores del coliambo y de los géneros satíricos y populares en general, a quienes hemos de atribuir la versificación de las fábulas. Las versificaciones, mejor dicho. Porque la Augustana contiene a veces restos de dos versiones poéticas de una misma fábula. Y conservamos restos de una colección de fábulas en el papiro Rylands 493 que contiene fábulas que también están en la Augustana, pero que aquí presentan restos de verso diferentes.

La fábula es un género popular que no fue cultivado por "fabulistas" autores de colecciones, como Fedro y Babrio, hasta muy tarde y en casos especiales. Todo el mundo se creía con derecho a modificar y retocar las fábulas anteriores. Los cínicos no hicieron más que esto, produciendo versiones diferentes, primero en verso y luego prosificadas, de unas mismas fábulas, y propagándolas en colecciones. La fábula es sólo uno entre los varios géneros de origen antiguo que los cínicos adoptaron para sus fines propagandísticos. Usaban por sistema la mezcla de lo burlesco y lo serio, de la sátira y el adoctrinamiento; si se apoderaron de la fábula para sus fines es porque este género popular y crítico coincidía en buena medida con sus ideas: preeminencia de lo "natural", ataque contra los poderosos brutales o hipócritas, contra la insensatez; elogio del pequeño inteligente y libre.

En este momento, hay que hacer un inciso para hablar de los cínicos, esa secta fundada por Diógenes de Sínope, el "Sócrates loco", que hizo furor en el siglo III a. C., sobre todo, y ejerció un vasto influjo fuera de la escuela estricta. El cínico no es un teórico: abomina de toda teoría, toda ciencia, toda cultura. Es un "sabio" que no soporta la idea del poder ni de las convenciones antinaturales como son las fronteras que dividen al mundo, las clases, los tabús sexuales, la falta de libertad, las necesidades consumistas, la moda. Es en cierto modo un anárquico, un asocial, que recorre el mundo con su manto basto y su alforja, burlándose de la sociedad y criticándola.

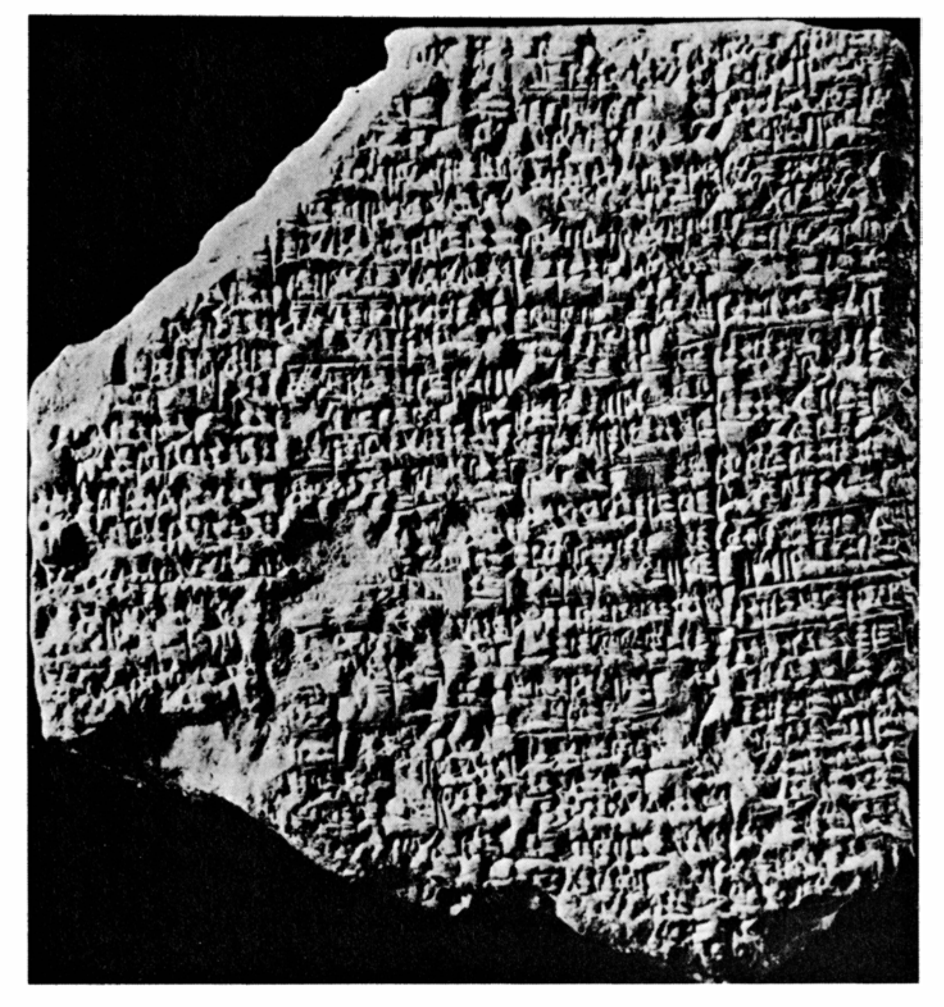
El dinero, el poder, la belleza, la insensatez, todo lo convencional, son sus verdaderos enemigos. Se contenta con poco y cree en las virtudes del trabajo, de la solidaridad humana. Critica a los adivinos, a los médicos, a los atletas, a las mujeres. Predica la completa libertad sexual (para los hombres solo, a decir verdad).

Este movimiento se apartaba del estado y del mundo oficial y se volvía al individuo. Representaba una disolución de la antigua sociedad, pero, al tiempo, una reafirmación de valores morales que influyeron fuertemente en el estoicismo e incluso en el cristianismo. Representaba, sobre todo, un estilo de vida y un estilo literario, también. La literatura era para él arma de combate, a base de punzar e irritar al hombre medio y conformista, de ofrecer también una lección.

Como decimos, la fábula fue utilizada por muy buenas razones por el movimiento cínico. Claro está que las antiguas fábulas recogidas por Demetrio fueron a veces retocadas, haciéndose entrar en ellas temas cínicos como el del rechazo de la *ánoia* o insensatez, que la zorra echa en cara al macho cabrío que, en el pozo, le ofrece ayuda para salir, pensando que luego va a sacarle a él. O como el de la *phusis* o naturaleza, la *tukhe* o fortuna, el aborrecimiento de la *truphé* o molicie, el elogio del *ponos* o esfuerzo.

Ahora bien, muy frecuentemente los cínicos inventaron nuevas fábulas, bien "fabulizando" motivos antiguos, bien desarrollando temas de las antiguas fábulas, bien creando libremente. Muchas de las fábulas que presentan en la Augustana y el papiro Rylands restos de yambos son fábulas cínicas (algunas de ellas pasaron también a Fedro y otros fabulistas). Y otras enteramente en prosa, entradas posteriormente (a partir del siglo II a. C.) en las colecciones, también. Ya nos presentan los viejos protagonistas de la fábula entre los que la zorra se convierte en un símbolo cínico, ya otros nuevos o usados en forma nueva. Son símbolos cínicos, entre otros, la tortuga que lleva la casa a cuestas y que con su esfuerzo vence a la rápida liebre; el mosquito que punza y vence al león y al atleta; la rana que asusta e irrita con sus gritos: la mosca imprudente; el caminante.

Hay que saber que la fábula cínica comprende muchas anécdotas, algunas con intervención de personajes divinos, como Hércules (el modelo del cínico),



FRAGMENTO DE TABLILLA DE LA SEGUNDA COLECCION SUMERIA de proverbios y fábulas. El contenido de unos y otras suele coincidir en Mesopotamia, Grecia y la India, en muchos ejemplos.

Prometeo, Hermes; otras con personajes humanos de sabios como Sócrates, Esopo o Semónides. Son frecuentes, también, las fábulas "de situación", con una circunstancia y un comentario, y aquellas en que el enfrentamiento se refiere a la disputa entre dos animales o plantas sobre cuál de los dos tiene cualidades preferibles. Entre las fábulas sobre el tema de la naturaleza, abundan aquellas en que el animal o animales que protestan van a Zeus o le envían una embajada, recibiendo la respuesta de que lo sabio es, precisamente, la naturaleza.

Los temas son varios. Contra la riqueza y la codicia los cínicos inventan fábulas como la de Hércules que, recibido en el Olimpo, no saluda a Pluto, la Riqueza, porque "siempre se le ve acompañado de los malos"; o como la de la mujer que mata a la oca de los huevos de oro o la zorra o el ratón que, habiendo comido demasiado, no pueden salir de su guarida o refugiarse en ella. A la belleza física del leopardo moteado contesta la zorra: "yo tengo moteada el alma": es la belleza espiritual. La belleza sola no basta: ¿cómo podrá el pavo real defender a las aves si es elegido rey? Y la máscara trágica es hermosa, pero no tiene seso. Tampoco es decisiva la fuerza bruta: el león es vencido por el mosquito, el atleta por la pulga.

El ideal está en aquel a quien sus riquezas y pasiones no ponen en cuidado, como el perro, que, exhortado por su amo a que se prepare para el viaje, responde: "Yo estoy preparado, eres tu el que se tarda". El perro es precisamente un símbolo cínico. El del hombre inteligente, que sabe manejarse en la vida al contrario que el cuervo de la fábula o que la máscara bella pero sin seso. Y que no se turba como el león asustado por el croar de las ranas: "No asustarse antes de ver", es la moraleja. En definitiva, el ideal es vivir la vida misma sin complicarse con el poder, el dinero o la ciencia, como el astrónomo que se cayó al pozo por contemplar las estrellas. El niño que se está ahogando dice al que le reprende desde la orilla: "Sálvame, ríñeme luego". El cínico, en definitiva, no tiene demasiadas exigencias en su vida: otro símbolo suyo es la mosca que se está ahogando en la sopa y que dice al morir: "He comido, he bebido, me he bañado: si me muero, ¿qué me importa?".

Otra serie de fábulas y novelitas cínicas es aquella en que hay burla o critica de los sabios, los médicos, los atletas, los adivinos, las mujeres: ya



COMADREJA CONVERTIDA EN MUJER, según ilustración de Grandville (siglo XIX). Representa el triunfo de la naturaleza. Este tema griego tiene su par (con la variante del animal) en la India.

hemos hablado de la hostilidad de la escuela (hay una excepción: Hiparquia, la mujer cínica que se casa con Crates y se pone el manto basto y la alforja para recorrer con él los caminos). En las colecciones de fábulas antiguas, por referirnos al último tipo, se encuentra una colección de historietas eróticas, maliciosas y a veces obscenas, que son sin duda alguna el precedente de la cuentística medieval.

Ahora bien, ya decíamos que en nuestras colecciones de fábulas griegas en prosa existen también fábulas que nunca contuvieron versos: fábulas cínicas, muy frecuentemente. Es que, en una fecha que podemos situar en los siglos I y II a. C., las fábulas en verso se prosificaron y entraron otras nuevas también en prosa. De otra parte, cuando el sentido de la fábula no estaba claro, se añadieron promitios o epimitios ocasionales, como sucede todavía en Fedro. Ciertas colecciones los generalizaron: la del papiro Rylands lleva siempre promitios, la Augustana siempre epimitios. Esta generalización de los epimitios debió de tener lugar en una fase antigua de la misma Augustana, probablemente en el siglo I a. C.



138 ÆSOPI VITA.

iuste eum interfecisse, etiam cippum erexerunt. Sed primates Græciæ ac sapientissimi quique, quum & ipsi, quæ in Æsopum facta fuissent, intellexissent, & in Delphos profecti sunt, & cum illis habita inquisitione, vltores & ipsi Æsopi mortis fuerunt.

Finis Æsopi vitæ à Maximo Planude conscripta.

ÆSO

139

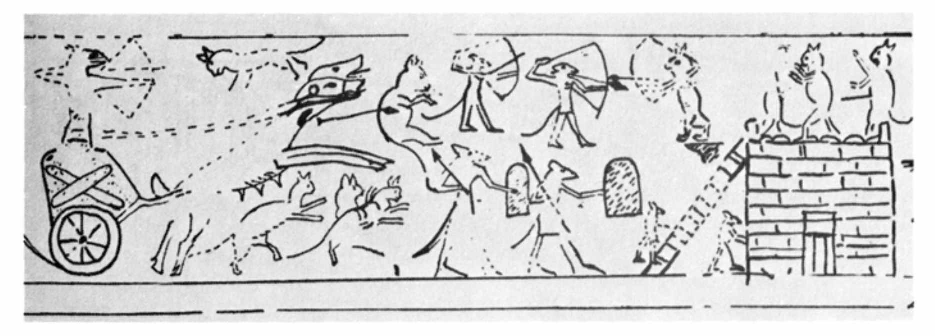
ΑΙΣΩΠΟΥ ΜΥΘΟΙ.

ÆSOPI FABULÆ.

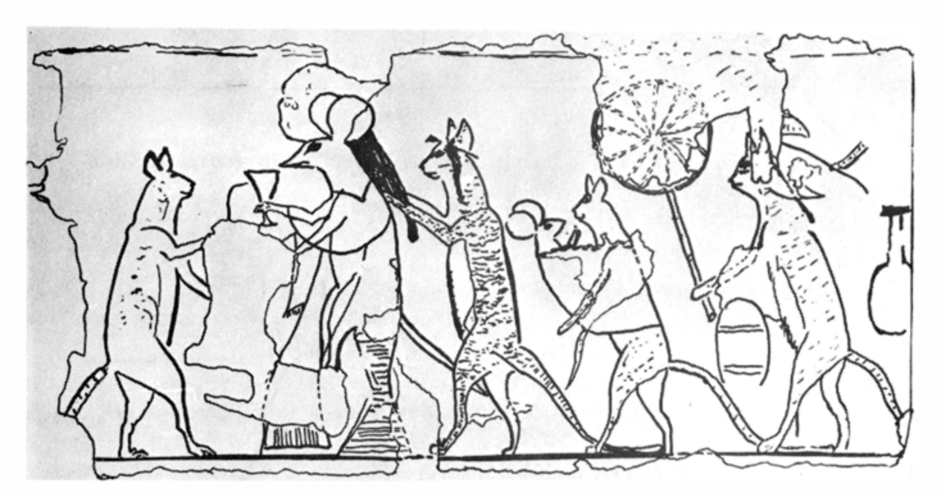
AQVILA ET Vulpes.

AQVILA & Vulpes inita amicitia, propè habitare decreuerunt, amicitiã familiaritate confirmãtes. Itaque Aquila super alta arbore

PAGINA FINAL de la *Vida de Esopo* e inicial de las fábulas. La fábula griega se irradió hacia el mundo oriental a través de traducciones siriacas. Esta biografía se tradujo al árabe, etíope, siriaco y turco.



LUCHA DE LOS GATOS Y LOS RATONES, según un papiro egipcio de Turín. No está suficientemente esclarecido, debido a la escasez de datos, el influjo real de la fábula egipcia en la griega helenística.



GATAS SIRVIENDO a una dama ratón. (En la primera parte de la historia la suerte les sonreía a los ratones.) Esta fábula de la guerra entre gatos y ratones penetró en Occidente a través de árabes y turcos.

Las colecciones de fábulas antiguas, griegas y latinas, que nos han llegado son escasas. La situación en la Antigüedad era muy diferente: había colecciones muy numerosas, que añadían, quitaban o contaminaban fábulas o daban nuevas versiones de las mismas. Estas colecciones pasaron en un momento dado a convertirse en material de enseñanza, un papel que la fábula no ha abandonado desde entonces hasta ahora. En realidad, nuestro más antiguo testigo de este nuevo uso de las fábulas es Quintiliano, en el s. I d. C. En las escuelas de retórica la fábula se usaba para la enseñanza de la redacción y la composición literaria, diríamos ahora: se prosificaban, alargaban, reducían.

Pero es que, al propio tiempo, las fábulas habían dejado de ser propiedad de los cínicos. Hacia el s. I a. C. hay que señalar la presencia en las fábulas de temas estoicos y moralistas en general: intervención de la providencia divina, castigo del malo. Las fábulas más groseras e indecentes fueron, al tiempo, eliminadas de varias colecciones. Así la fábula a través de los socráticos y los cínicos siguió un camino que la hacía útil para la expresión de estoicismo y el moralismo en general y más tarde, en la Edad Media, del cristianismo.

Pero hay que notar que esto no fue siempre así. En la época de Tiberio, Fedro, un liberto griego (macedonio en realidad) escribió su colección de fábulas, fuertemente cinizante y cuya intención no era en absoluto didáctica. Fedro, nuevo Arquíloco, utiliza la fábula como arma de ataque, de sátira encubierta del ambiente general de la corte y del favorito Seyano. Depende de la tradición de la fábula griega: su modelo es la antigua Augustana que luego evolucionó en griego independientemente. Pero ese modelo prosificado Fedro vuelve a versificarlo y añade toques personales muy acordes con las escuelas cínica y estoica y con el antiguo carácter satírico del género.

Otra colección importante, esta griega, es la de Babrio, que se fecha tentativamente a comienzos del siglo II d. C. Está escrita en verso, en coliambos: pero coliambos mucho menos laxos, más elaborados que los usados luego en fecha helenística por los cínicos. Babrio, concretamente, se jacta de haber renovado el coliambo y el género fabulístico. Tenemos razones para pensar que depende directamente de las colecciones coliámbicas helenísticas, no de las prosificaciones posteriores. La intención de la colección es puramente literaria: Babrio cultiva el arte de la narración fabulística, como muchos siglos más tarde La Fontaine. Aunque hay que hacer notar que Babrio tuvo muchos imitadores y que en la colección que hoy conocemos como de Babrio hay, seguramente, fábulas ajenas a él.

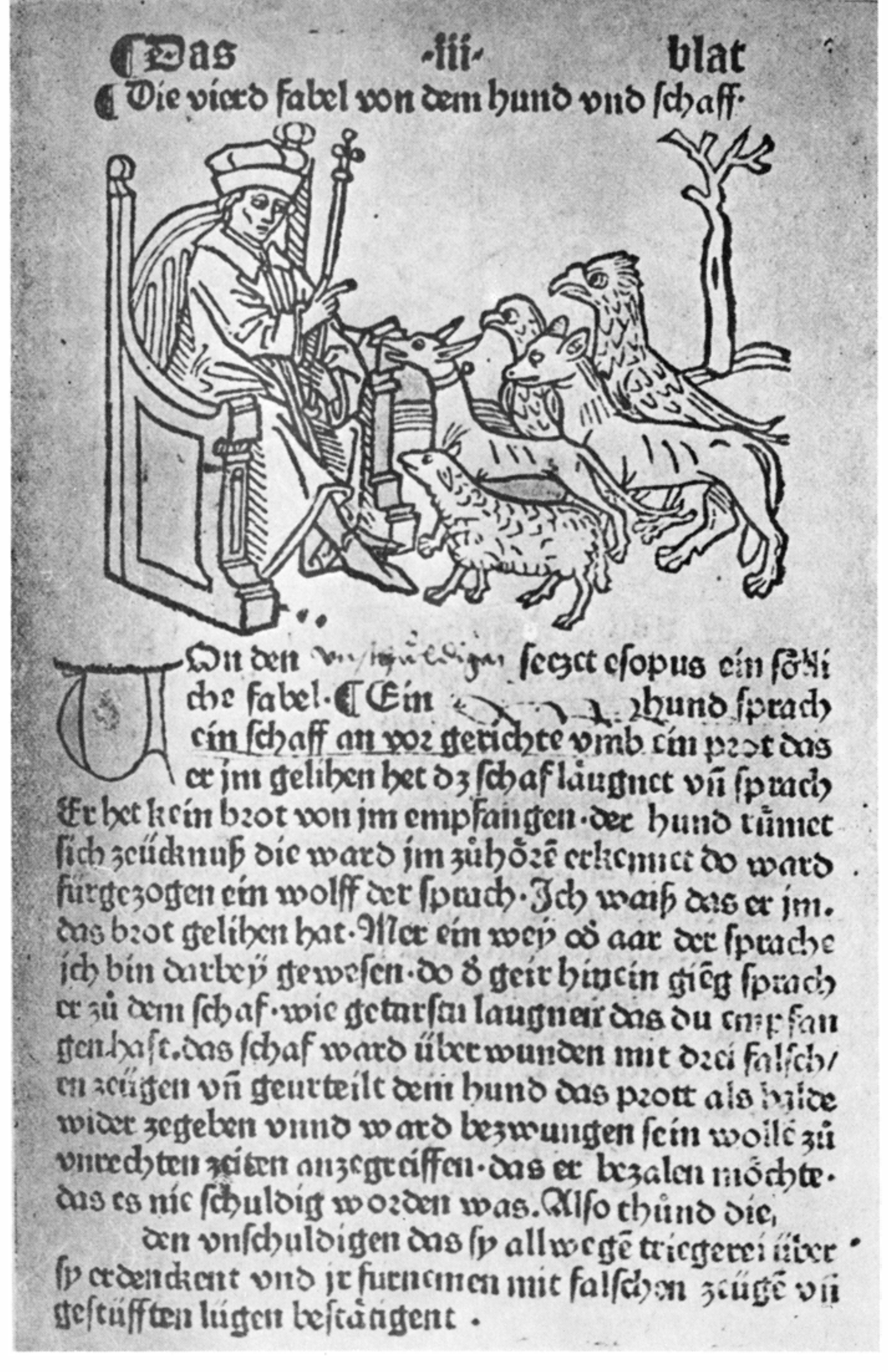
Con esto no está terminado, ni mucho menos, el inventario de la fábula de la época imperial. Hemos aludido al fabulista latino Aviano, muy emparentado con Babrio; y hay fabulistas griegos diversos, como Aftonio, el retor del códice Brancacciano, el llamado Pseudo-Dositeo, etc. Comparando estas fuentes y otras más, así las fábulas citadas como ejemplo por tantos autores (Horacio, Luciano, Plutarco, etc.), se puede investigar la complicada historia de la fábula en esta época.

Habría que añadir, para esta investigación, otras fuentes. Así las fábulas latinas, medievales, de Rómulo, que

DOS PAGINAS del manuscrito G de la *Vida de Esopo*, datado en el siglo X. En Bizancio era muy popular, sobre todo en la redacción de Westermann (que se copiaba encabezando las fábulas de la Vindobonense).

junto a elementos derivados de Fedro contienen otros procedentes de colecciones perdidas; fábulas conservadas en siriaco (y de ahí traducidas al árabe) y que son de origen griego, figurando entre ellas algunas de las que no se nos han conservado versiones griegas; las fábulas del llamado Sintipas, que son antiguas fábulas griegas traducidas al siriaco y de ahí, de nuevo, al griego en el s. XI. Todavía hay que añadir fábulas aisladas conservadas por diversos papiros y las de las llamadas Tablas de Asendelft, tablas enceradas en que un niño de Palmira copiaba en el s. III d. C. sus ejercicios de clase.

Muy interesante también para el estudio de la fábula en la antigüedad greco-romana, y también en la India, es la relación que se puede encontrar entre las colecciones indias y las griegas allí donde esta relación no es atribuible al común influjo de Mesopotamia. Hay, efectivamente, razones para pensar que diversas fábulas del *Tantrākhyāyika* y del *Pañchatantra* contienen elementos procedentes de la fábula griega de tipo cínico. Por ejemplo, la fábula del perro codicioso que perdió la carne que llevaba cuando, por error, confundió su imagen en el río con otro perro que llevaba carne: al querer arrebatársela, perdió la carne propia. En la India el perro pierde la carne al querer pescar un pez; y existe la fábula del león que se ahoga al querer luchar con otro león que, en realidad, es su propia imagen reflejada en un pozo.


Das ·iii· blat
Die vierd fabel von dem hund vnd schaff·

Von den [...] seczet esopus ein söli che fabel· Ein [...] hund sprach ein schaff an vor gerichte vmb ein prot das er im gelihen het dz schaf laugnet vñ sprach Er het kein brot von im empfangen·der hund rümet sich zeücknuß die ward im zůhörē erkennet do ward fürgezogen ein wolff der sprach· Ich waiß das er im. das brot gelihen hat· Aber ein wey oď aar der sprache ich bin darbeÿ gewesen·do ď geir hinein gieg sprach er zů dem schaf·wie getarstu laugnen das du emp fan gen hast. das schaf ward überwunden mit drei falsch/ en zeügen vñ geurteilt dem hund das prott als balde wider zegeben vnnd ward bezwungen sein wolle zů vnrechten zeiten anzegreiffen·das er bezalen möchte· das es nie schuldig worden was. Also thůnd die, den vnschuldigen das sy allwegē triegerei über sy erdenckent vnd jr furnemen mit falschen zeügē vñ gestüfften lügen bestätigent ·

EL PERRO Y LAS OVEJAS, fábula incluida en la edición de Steinhöwel (posterior a 1475). En la segunda mitad del siglo XV lo fundamental de la fábula griega volvió a ser conocido en Occidente.

Pero no se trata tanto de fábulas particulares como de la presencia en la fábula india de temas cínicos sobre la naturaleza, la riqueza, etc., de promitios y epimitios semejantes a los griegos; y de que la propia idea de crear una colección de fábulas difícilmente apareció en la India sin el modelo de las fábulas griegas. El *Tantrākhyāyika* (y su derivado el *Pañchatantra*) es en realidad un tratado para la educación de príncipes en el gobierno del estado, que combina el tema mesopotámico de la instrucción que imparte un padre o un filósofo con la idea griega de la colección de fábulas. Dado que se trata de fábulas en prosa que contienen promitios y epimitios, hay que situar su origen, probablemente, en el s. II a.C., época de las colecciones prosificadas griegas.

Es que una serie de tendencias del pensamiento y la religión india, que caían más bien fuera de la tradición brahmánica, coincidían bien con las de los cínicos y las de la fábula griega helenística en general. Para los griegos, los santones desnudos indios eran verdaderos cínicos por su ideal de vida simple y ascética. De otra parte, es claro que los griegos establecidos en la India se sintieron especialmente próximos al budismo y a las ideas humanitarias del rey Aśoka, en el s.III a. C.: ideas de compasión, de consideración del animal como digno de respeto, de desprecio de la riqueza y pensamiento en la muerte.

De esta manera, la fábula mesopotámica ha pervivido en cierto modo a través de las dos ramas de la fábula india y griega. Y, más tarde, ambas ramas se han contaminado, por el influjo griego sobre la fábula india, seguramente como decimos en el s. II, época de los reinos indo-griegos de Demetrio y Menandro. Pero al tiempo la fábula griega irradiaba, ya lo hemos dicho, hacia el mundo oriental a través de traducciones siriacas que encontraban difusión en otras lenguas: ni más ni menos que como la *Vida de Esopo* se tradujo al árabe, el etíope, el siriaco y hasta el turco. Y habría que añadir otro tema sobre el que tenemos pocos datos, aunque curiosos: el influjo de la fábula egipcia, que conocemos, aunque

mal, desde el s. XIV a. C. y que ha dejado huella en algunas fábulas de las colecciones griegas helenísticas (y quizás en otras más antiguas). Y que en algún caso, así el de la guerra de gatos y ratones, ha penetrado a través de árabes y turcos en Occidente.

Al comienzo de la Edad Media, la fábula aparece dividida en tres grandes espacios. En el oriental, bizantino, continúa la existencia de la fábula griega. La Augustana es la colección griega prosaica, continuadora de Demetrio, que adquirió su forma definitiva en el s. V d. C. y fue fielmente copiada desde entonces. Pero produjo dos derivados que contienen fábulas con versiones modificadas: la colección Vindobonense, del s. VII, popular y aún vulgar; y la Accursiana, del s. IX. Esta es una manera de hablar porque cada colección se descompone en variantes infinitas.

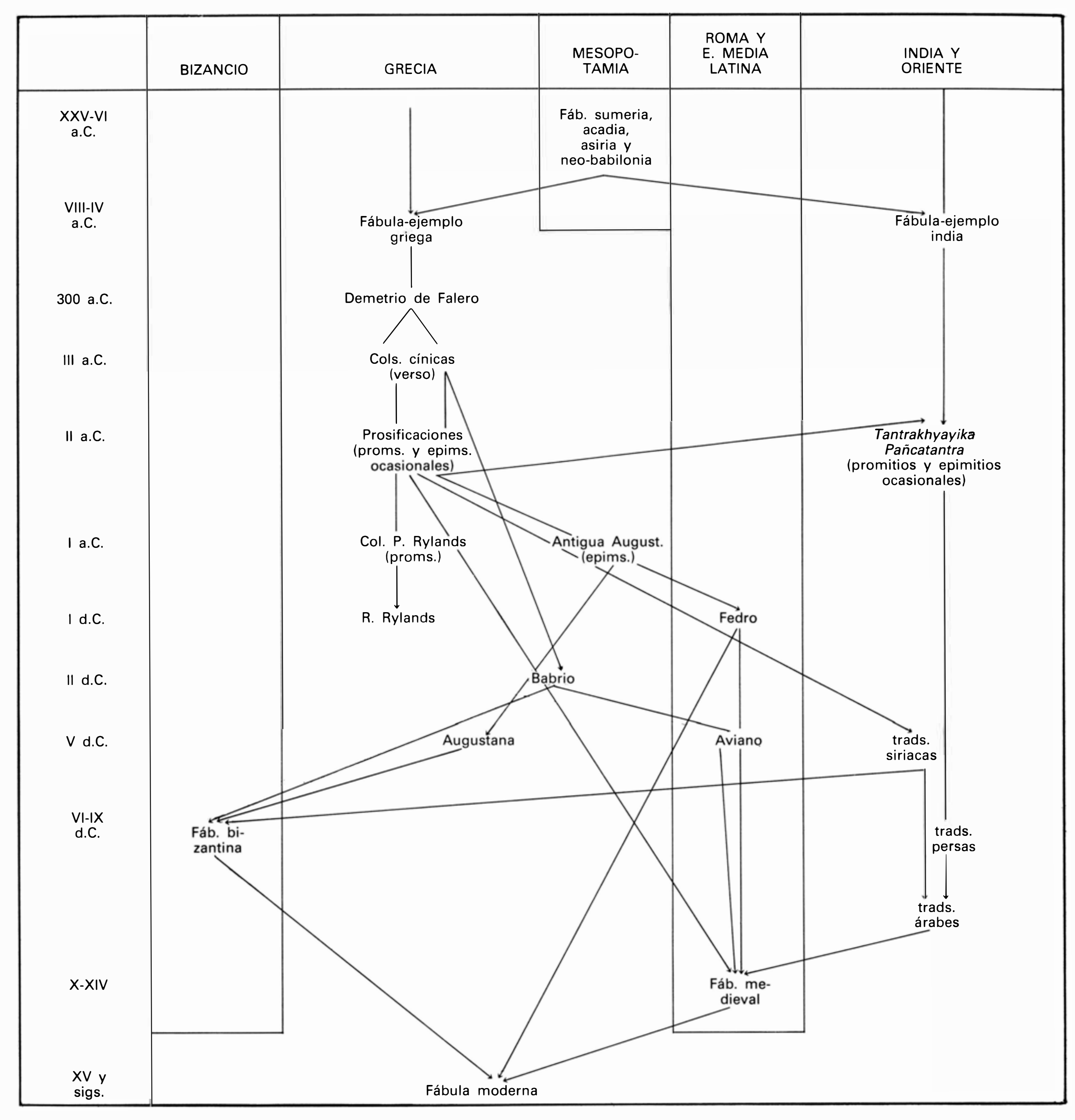
Junto a esta tradición existió en Bizancio la derivada de Babrio y sus imitadores. Apenas fueron copiados, pero en cambio se hicieron numerosas prosificaciones o "paráfrasis", que gozaron de gran popularidad; y éstas prosificaciones, a su vez, se versificaron de nuevo a veces, ya por obra de versificadores anónimos que usaban el verso "político" bizantino, de ritmo acentual, ya por otros más tradicionales como Ignacio Diácono (s. IX). De otra parte, en Bizancio era muy popular la *Vida de Esopo,* sobre todo en la redacción llamada de Westermann (que se copiaba encabezando las fábulas de la Vindobonense) y la de Planudes, del s. XIV (copiada encabezando la Augustana). Las fábulas de esta *Vida* pasaban a veces a las colecciones.

El segundo ámbito es el de la Europa occidental. Aquí Fedro fue poco leído, pero se hicieron paráfrasis en prosa muy divulgadas y se propagó mucho también el llamado Rómulo, que contiene fábulas en prosa derivadas ya de Fedro ya de colecciones perdidas. Rómulo fue a su vez modificado por varios escritores latinos, de entre los cuales el más conocido es Walter el inglés (modelo de nuestro Arcipreste). Hay, de otra parte, fábulas latinas diversas cuya genealogía es, hoy por hoy, difícil de trazar.

Y con esto llegamos al tercer ámbito, el oriental. Aquí se inicia un camino que llevará más tarde a la síntesis de las tres tradiciones, herederas en definitiva de la de Mesopotamia. Este camino es el de la traducción de las fábulas indias al pehlví en Persia y, luego, al árabe. Fue en la corte de Cosroes el Grande, en el



EL GATO PREDICANDO A LAS GALLINAS. **Las sillerías de los coros catedralicios (la foto es un detalle de la de Zamora, s. XVI) y monacales testimonian la importancia de este género literario.**



DESDE LOS ANTIGUOS SUMERIOS, hacia el año 2500 a. C., hasta nuestros días, la fábula se ha mantenido con rasgos muy fijos como un género popular y crítico. Podemos reconstruir su historia a través de ramificaciones y de confluencias dentro de una tradición fundamentalmente unitaria.

imperio sasánida, donde hacia el año 570 el médico Burzuyeh tradujo el *Pañchatantra* del sánscrito al pehlví; de aquí proviene la traducción árabe de al-Mukaffa, realizada en Bagdad, en la época de al-Mansur (754-775) (y también una traducción siriaca más antigua, del s. VI). Es la fuente de nuestro *Calila e Dimna* medieval (de 1251) y de infinitas versiones, directas o indirectas, a numerosísimas lenguas, incluso al latín, a fines del s. XIII, por Juan de Capua (a partir de una versión hebrea). De otra parte, obras medievales como el *Sendebar* (cuyo texto castellano, de 1253, se titula *Libro de los engaños e asayamientos de las mugeres*) y *Las Mil y Una Noches* han recorrido, en definitiva, igual camino a partir de la India.

La síntesis de estos ámbitos fabulísticos se produjo en la Europa occidental a partir del s. X, como se ha dicho más arriba. Transcurrió en dos fases: la primera, fusión de la tradición india con la latina de origen griego. En realidad hay que distinguir tres líneas principales:

1. Obras en latín, como la *Disciplina Clericalis* de Pedro Alfonso (s. XII), o en castellano, como el *Conde Lucanor* de D. Juan Manuel o el *Libro del Buen Amor* de nuestro Arcipreste de Hita (ambos del s. XIV) en que se toma el marco o armazón de la tradición oriental: instrucción o diálogo a base de fábulas. Ahora bien, esas fábulas son ya de la tradición oriental, ya de la latina indistintamente. Y, siguiéndose en esto la libertad que desde antiguo procuraba el género, se introducen a veces fábulas o anécdotas contemporáneas; esto es visible sobre todo en el

Conde Lucanor. Por otra parte, este tipo de composición es el modelo de obras como los *Cuentos de Canterbury* de Chaucer o el *Decamerón,* de Boccaccio. El que aquí se trate de anécdotas y novelitas no tiene por qué chocar: las hay desde siempre en las colecciones de fábulas greco-latinas y también en las de la India, seguramente por influjo griego en época helenística. O quizá sea tradición más antigua todavía; ya en Mesopotamia aparecen novelitas junto a las fábulas y máximas.

2. Otra línea, más antigua que la anterior, está formada por obras latinas como la *Ecbasis Captivi* (s. x) y el *Ysengrimus* (s. XII), poemas animales europeos en verso. El primero se localiza en Lorena, en un ambiente monástico; el segundo, en Flandes. Aquí hay que considerar que el marco es igualmente oriental: se trata de una fábula animal despaciosamente narrada, como en la India, y que incluye dentro otras, también como en la India. A la India remontan, aparte de fábulas aisladas que se incluyen, temas como el de la corte del león con sus distintos dignatarios y el hecho mismo de que los personajes animales tengan nombre propio: Nobilis es el león, Ysengrimus el lobo, Reinardus la zorra, etc. Ahora bien, la mayor parte de las fábulas que se narran son de tradición greco-latina, aunque las hay inventadas y otras con motivos antiguos modificados en el nuevo ambiente cristiano. Esta línea es la que produce, en Francia, a partir del s. XII, el *Roman de Renart,* dividido en distintas *branches* o episodios que relatan las diversas aventuras de la zorra. Esta obra, traducida al alemán y otras lenguas, tuvo una gran difusión. En ella la fábula vuelve a cobrar su antigua función crítica, esta vez en relación con el clero y la nobleza.

3. Finalmente, no faltan en la Edad Media latina fábulas aisladas o en colección que son parcialmente nuevas. Entre ellas son notables las que critican la hipocresía de cierto clero, representado bajo la figura del lobo o el gato que no pierden sus mañas aunque entren en el convento. Aunque pueden inspirarse en modelos ya indios (sátira de los brahmanes) ya greco-latinos (tema del gato médico y otros emparentados). A partir del s. XIV este sincretismo de tradiciones que tenía lugar en la Europa occidental se incrementó con la entrada a partir de Bizancio de la tradición propiamente griega que allí se había conservado.

Así, a partir de la traducción latina de la *Vida de Esopo* (por Rinuccio Aretino, 1446-1448), de la inclusión de fábulas griegas en la edición de Steinhöwel (latín y alemán, posterior a 1475) y de la edición de la colección Accursiana (1479-1480), lo fundamental de la fábula griega volvió a ser conocido en Occidente, aunque quedara para el siglo XIX el redescubrimiento de la Augustana y de Babrio. Toda esta literatura influyó grandemente en Europa. He sostenido en otra parte que el *Lazarillo* y toda nuestro novela picaresca debe mucho a la *Vida de Esopo,* que conoció en España diversas ediciones en traducción a partir de 1489. Y surgió más tarde la fábula como género literario independiente, en colecciones del tipo de las griegas, por obra sobre todo de La Fontaine en Francia en el siglo XVII, e Iriarte y Samaniego en España en el XVIII. Esta literatura, muy abundante, es en realidad poco conocida. El influjo de la fábula continúa hasta hoy.

Este influjo ha rebasado lo puramente literario. "Paciencias" de las sillerías de nuestros coros, artesonados como el de Silos, tapicerías como las de Bayeux testimonian la importancia de la fábula de la Edad Media en adelante. La fábula no se tomaba como algo frívolo, sino como enseñanza: todavía Felipe IV llamaba Esopo y Menipo a los dos bufones pintados por Velázquez, veía en el fabulista un personaje comparable en cierto modo al cínico. En la enseñanza elemental, de otra parte, continuaba el papel ya antiguo de la fábula: Fedro ha dominado durante mucho tiempo la enseñanza del latín.

Por otra parte, la fábula india, de un lado, y la greco-latina, de otro, se han extendido a todas las culturas de la tierra. La primera ha penetrado desde la Edad Media en Indonesia, el Turquestán, Mongolia, Africa; la segunda ha llegado a todos los lugares donde han puesto pie los europeos. En México, por ejemplo, se empezó a partir de la Conquista a escribir fábulas en que los animales indígenas sustituyen, a veces, a los tradicionales: las conservamos escritas en nahuatl. Lo notable es considerar que la fábula, que en sus ramificaciones se ha convertido en parte de la cultura universal, en la que se han confundido las distintas ramas, procede en último término de una fuente única: la antigua Mesopotamia, que ha ido absorbiendo las distintas ramas indígenas, entre ellas la griega y la india. Esto explica, pensamos, la unidad esencial de un género popular y crítico, un género menor tal vez, pero importante por su constancia y su adaptabilidad para expresar y criticar las diversas culturas e ideologías.

Galileo observa Neptuno

Lo observó por primera vez en 1612 y creyó que se trataba de una estrella fija, unos 234 años antes de que se descubriera que era un planeta. Sus datos hacen surgir dudas acerca de la exactitud de los cálculos modernos sobre su órbita

Stillman Drake y Charles T. Kowal

Galileo Galilei dirigió su telescopio hacia el firmamento por vez primera en 1609. Vio las montañas y valles de la Luna y, en el siguiente año, descubrió los satélites de Júpiter, la curiosa forma de Saturno y las fases de Venus. Hemos encontrado ahora, en uno de sus diarios, los informes acerca de otra observación que puede resultar útil para los astrónomos modernos: la primera observación de Neptuno. Aunque no lo identificó como un planeta, lo vio por primera vez en diciembre de 1612, unos 234 años antes de que fuera reconocido como el octavo planeta por Johann Gottfried Galle, un joven astrónomo de Berlín.

Nuestro examen del diario de Galileo trajo consigo otro resultado inesperado: un año, aproximadamente, antes de su observación de Neptuno, desarrolló un ingenioso instrumento de medida para determinar el tamaño aparente de los objetos celestes. A veces se ha partido de la base de que Galileo no podía utilizar un micrómetro con su telescopio y que, por consiguiente, fue incapaz de realizar observaciones precisas. Nuestro trabajo muestra, por el contrario, que sus observaciones eran de una precisión notable ya en el segundo año de la astronomía telescópica.

La fiabilidad de las observaciones de Galileo hace que las que realizó de Neptuno sean mucho más que una curiosidad histórica. Estas observaciones cuestionan la exactitud de la órbita de Neptuno tal como se ha calculado modernamente. Conocemos con gran precisión las órbitas de los primeros siete planetas; de hecho, la exactitud con la que podemos calcular posiciones planetarias ha logrado que la mecánica celeste se convirtiera en el paradigma de las ciencias exactas. Puede, por tanto, constituir una sorpresa el hecho de que la órbita del octavo planeta deba, tal vez, ser recalculada al tener en cuenta observaciones realizadas, con instrumentos primitivos, hace más de tres siglos y medio. Una de las razones que motivan las incertidumbres en la órbita de Neptuno es el largo período de revolución del planeta, que dura 165 años; desde su descubrimiento como tal planeta, en 1846, no ha completado aún una revolución en torno al Sol. Incluso si tenemos en cuenta esto, la órbita que se deduce de las observaciones realizadas desde 1846 debiera ser suficientemente precisa; las discrepancias entre las posiciones teóricas de Neptuno en 1612 y 1613 y las que registra Galileo no han recibido aún una explicación convincente. Una posible explicación es que un planeta por descubrir todavía esté perturbando la órbita de Neptuno.

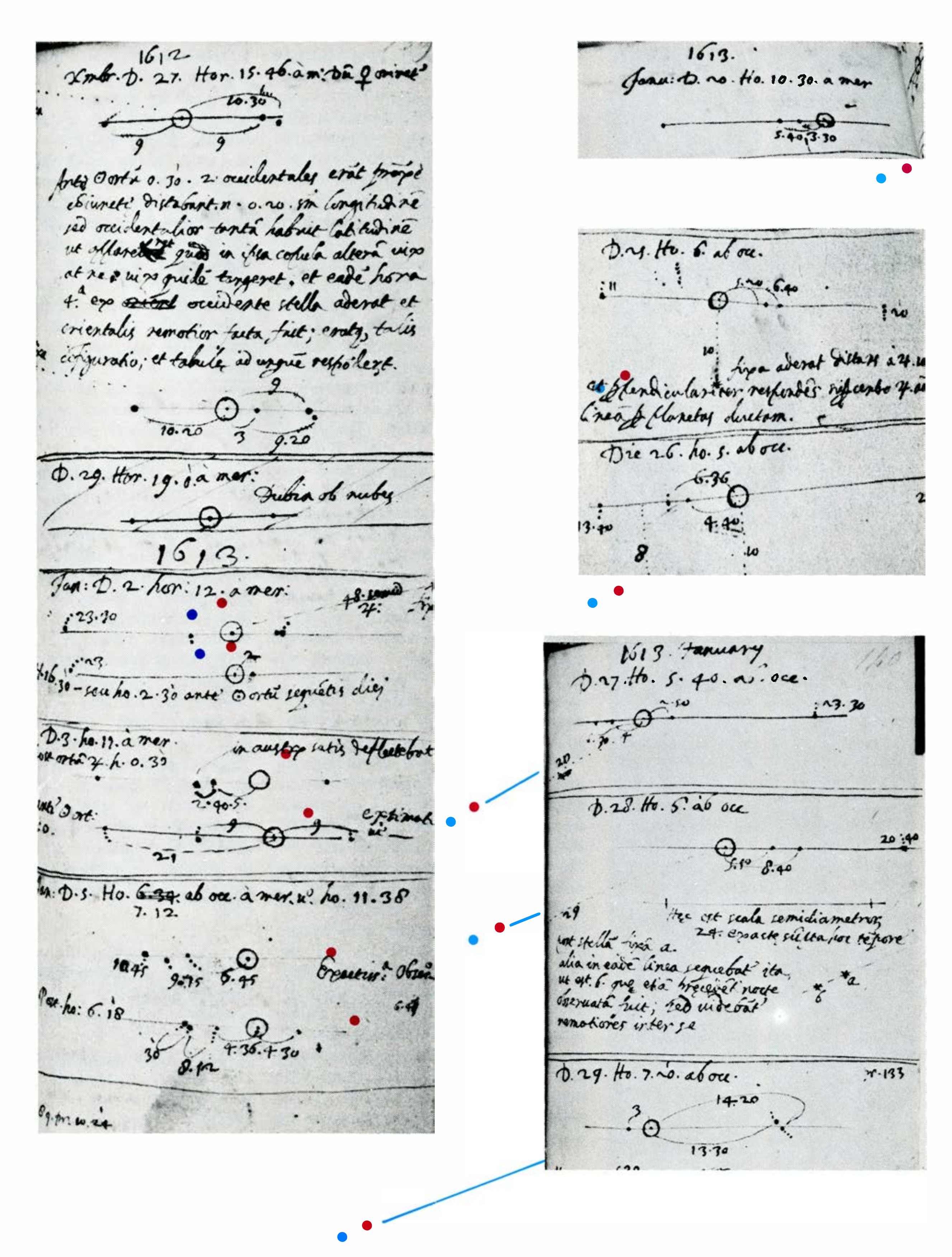
Averiguamos que Galileo había observado Neptuno como consecuencia de un intento de afinar los cálculos de la órbita de este planeta. Una observación primitiva tiene interés en tentativas de esta índole, ya que cualquier error en la órbita adquirirá mayor magnitud y evidencia con el transcurso del tiempo. Conocíamos ya una observación de Neptuno anterior a Galle. Joseph Lalande, un astrónomo francés que catalogó unas 50.000 estrellas, determinó su posición en 1795, si bien supuso que se trataba de una estrella fija. La posición dada por Lalande se encuentra a siete segundos de arco, o sea, a un quinientosavo de grado, al oeste de la posición calculada retrospectivamente. Puede que la discrepancia parezca pequeña, pero la astronomía telescópica era ya una ciencia exacta a fines del siglo XVIII y la diferencia no puede atribuirse, sin más, a un error de observación.

Para evaluar mejor la precisión de la órbita de Neptuno, uno de nosotros (Kowal) tenía la esperanza de poder descubrir observaciones primitivas del planeta. En 1979 Steven C. Albers publicó una lista de todas las ocultaciones de un planeta por otro, vistas desde la Tierra, de 1557 a 2230. En este período de casi 700 años se producen sólo 21 acontecimientos de esta índole y el tercero de la lista fue una ocultación de Neptuno por Júpiter en enero de 1613. Galileo observó metódicamente los satélites de Júpiter en esta época y podía haber visto Neptuno en el momento aproximado de dicha ocultación, pero indudablemente lo habría considerado como una débil estrella fija. Esta hipótesis pudo comprobarse: Galileo vio efectivamente Neptuno según ha revelado el examen de su diario.

Al igual que Lalande, Galileo señaló posiciones de Neptuno que se encuentran al oeste de las que cabría esperar al calcular la órbita teórica del planeta. La discrepancia puede alcanzar el valor de un minuto de arco o un sesentavo de grado. Hemos llegado a la conclusión de que la mayor parte de esta diferencia no debe atribuirse al carácter primitivo del instrumental utilizado o a errores de observación achacables a Galileo, sino a la inexactitud de la órbita calculada.

Puede resultar extraño el que tengamos en tan alta estima observaciones realizadas en los primeros años de la astronomía telescópica. Los motivos que nos mueven a ello quedarán patentes cuando discutamos la forma y el contenido del diario de Galileo, sus métodos de medición, las propiedades de su telescopio y la naturaleza de algunas de sus observaciones de satélites, así como sus tablas, cuya precisión puede determinarse si las comparamos con datos modernos.

En abril de 1611 Galileo había computado tablas de los movimientos de los cuatro satélites de Júpiter más brillantes (Io, Europa, Ganimedes y Calixto), a los que hoy en día denominamos satélites galileanos. En el diario en el que se registran las observaciones



APUNTACIONES en el libro de notas de Galileo para algunas noches seleccionadas correspondientes a diciembre de 1612 y enero de 1613; dichas anotaciones incluyen diagramas de Júpiter y de sus cuatro satélites de mayor tamaño. En cada diagrama, Júpiter se encuentra en el centro dejando, a cada lado, una distancia que representa 24 radios jovianos. Los cuatro satélites (Io, Europa, Ganimedes y Calixto) aparecen representados mediante puntos situados a lo largo de la línea horizontal y se distinguen unos de otros por puntos suplementarios que se colocan encima o debajo de la línea. Io no lleva ningún punto adicional, Europa uno, Ganimedes dos y Calixto tres. Las distancias entre Júpiter y un satélite, o entre dos satélites, se señalan utilizando un radio de Júpiter como unidad. El 28 de diciembre de 1612 y el 28 de enero de 1613, Galileo observó Neptuno, pero creyó que se trataba de una estrella fija. Sus datos muestran que Neptuno se encontraba, quizás, un minuto de arco al oeste del lugar que le correspondería de acuerdo con la órbita teórica moderna. Los autores han calculado, a partir de estos datos, dónde se hallaría Neptuno (*puntos rojos*) en las noches en las que Galileo no señaló su presencia. Los puntos azules dan su posición de acuerdo con la órbita teórica que corrientemente se acepta. Las apuntaciones del libro de notas, proporcionadas por la Biblioteca Central Nacional de Florencia (Italia), se reproducen a un tamaño igual al 80 por ciento del original.



DOS TELESCOPIOS de los que se cree que fueron construidos por Galileo y que se conservan en el Instituto y Museo de Historia de la Ciencia de Florencia. El instrumento de la izquierda tiene catorce aumentos, un poder de resolución de 20 segundos de arco y un campo visual de 17 minutos de arco. El instrumento de la derecha tiene 20 aumentos, un poder de resolución de 10 segundos de arco y un campo de visión de 17 minutos de arco. El instrumento que fue efectivamente utilizado por Galileo en sus observaciones de Neptuno, y que no conservamos, tenía 18 aumentos, un poder de resolución de 10 seg. de arco y su campo de visión era probablemente de 17 minutos de arco.

de Neptuno sólo anotó observaciones que pudieran ayudarle a perfeccionar sus tablas. El diario, que se conserva en la Biblioteca Central Nacional de Florencia, es un estrecho volumen que mide 29 × 9,5 centímetros. Incluye numerosos diagramas del sistema de Júpiter, la mayor parte de los cuales tienen el mismo formato. De manera característica, Júpiter se encuentra en el centro de la página y, a cada lado, dispone de una distancia equivalente a 24 radios de Júpiter. Cada margen deja espacio para otros 10 radios de Júpiter. Sin tener en cuenta la orientación real del plano de los satélites, Galileo lo dibujó como una línea horizontal con el este hacia la izquierda. Las posiciones de los cuatro satélites se marcan mediante puntos situados sobre la línea; suelen distinguirse unos de otros mediante un sistema de puntos suplementarios situados por encima o por debajo de la línea. Io, el satélite más interno, carece de puntos adicionales de identificación, Europa lleva uno, Ganimedes dos y Calixto tres. La elongación de un satélite (su distancia aparente desde el centro de Júpiter) se expresa utilizando como unidad el radio del planeta. En algunas ocasiones Galileo definió la escala del diagrama citando la distancia entre dos satélites en lugar de la elongación del satélite más alejado.

Mucho antes de que se inventara el telescopio, los cálculos astronómicos se aproximaban no sólo hasta los minutos de arco, sino incluso hasta pequeñas fracciones de un segundo de arco. No obstante, prescindiendo de la precisión numérica de tales cálculos, las observaciones astronómicas resultaban rara vez más exactas que un minuto de arco y a veces incluso menos. Tycho Brahe que fue, quizás, el observador más preciso de la era anterior a la introducción de un sistema de lentes, estimó que el diámetro aparente de Júpiter y los de las estrellas brillantes llegaban a alcanzar hasta dos y tres minutos de arco. Galileo se dio cuenta de que las estimaciones de Brahe incurrían en errores de hasta un 3000 por ciento.

¿Cómo logró Galileo realizar mediciones de ángulos pequeños y obtener valores que aún hoy están en uso? Su método para utilizar el telescopio como instrumento de medida no ha sido conocido durante mucho tiempo, al no publicar él gran cosa sobre el tema. Suele suponerse, además, que las mediciones astronómicas precisas empezaron sólo con el micrómetro con retículo capilar que se perfeccionó tras la muerte de Galileo. Un micrómetro de esta índole no podía utilizarse con el telescopio de Galileo, que, al igual que los gemelos de teatro, genera una imagen virtual. El micrómetro con retículo capilar requiere una imagen real como la que proporciona el telescopio kepleriano o astronómico. No obstante, Galileo fue capaz dé desarrollar un tipo de instrumento micrométrico que funcionaba razonablemente bien con su telescopio.

El último día de enero de 1612, cuando Galileo había medido el diámetro de Júpiter que estimó en 41 segundos de arco, escribió en su diario que había empezado a utilizar "un instrumento para medir intervalos y distancias de manera exacta, aunque el artilugio en cuestión no está construido aún de modo muy preciso". Aunque no describió el instrumento, que no se ha conservado, su discípulo Giovanni Alfonso Borelli lo mencionó en 1666.

Según Borelli, el instrumento micrométrico era un retículo trazado con toda precisión al que Galileo superpuso ópticamente la imagen telescópica. La idea puede resultar poco plausible pero, de hecho, funciona bastante bien. Invitamos al lector a ensayar una técnica similar. Tómese una hoja de papel milimetrado y sitúese cerca del extremo de un pequeño telescopio: mírese al papel con el ojo izquierdo mientras se observa con el otro a través del telescopio. Resulta fácil ver que la imagen amplificada se superpone al papel. Resulta obvio, de manera inmediata, el procedimiento por el cual, con un poco de esfuerzo, se podían realizar mediciones precisas utilizando este principio.

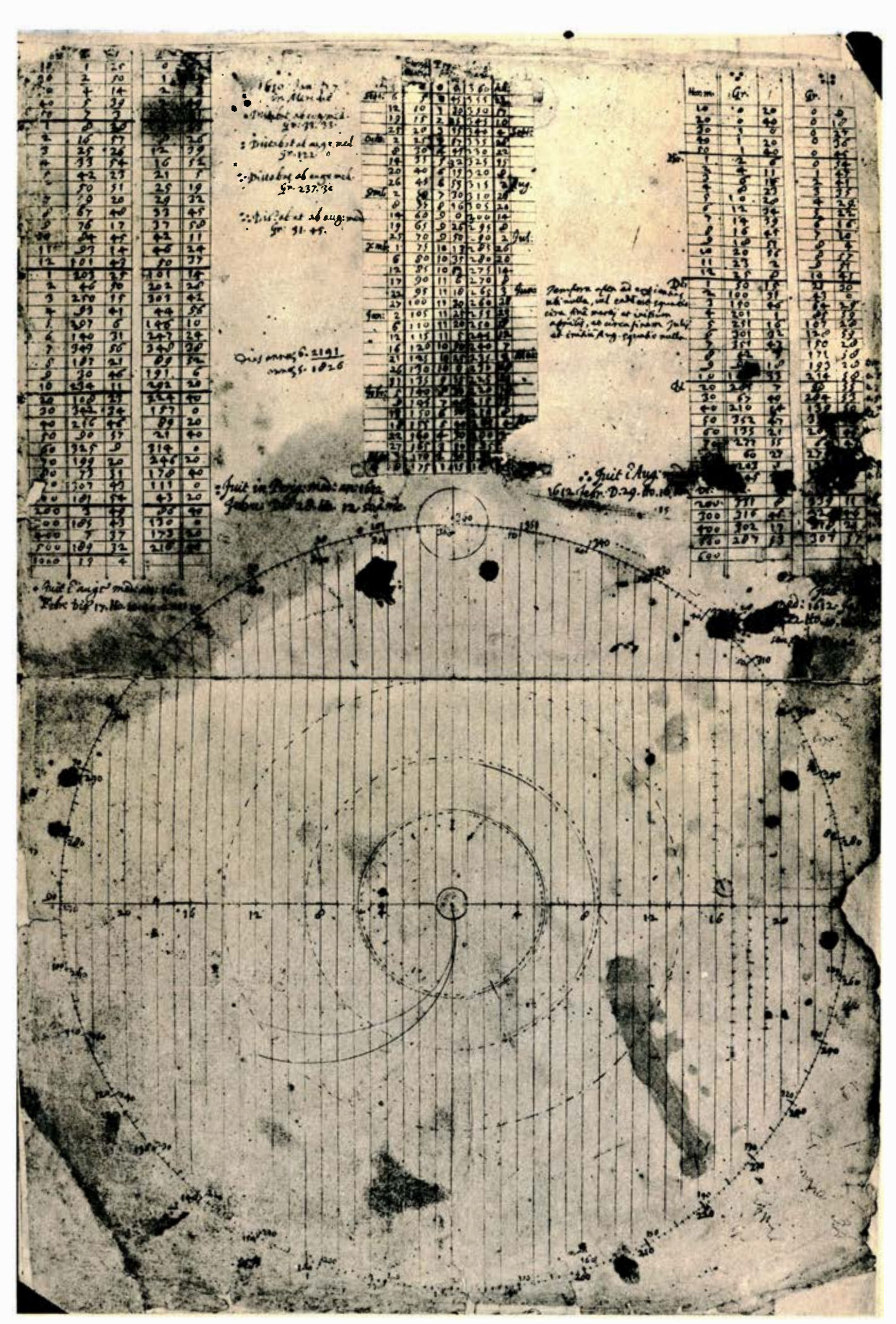
Ni Galileo ni Borelli explicaron con detalle la construcción del instrumento micrométrico. No obstante, basándonos en varios comentarios de Galileo y en la comparación que hemos realizado entre muchas de sus mediciones y cálculos modernos, estamos bastante seguros de que el instrumento se construyó sobre las bases siguientes. Había un retículo circular con un diámetro de diez centímetros, y un espacio de unos dos milímetros separaba cada subdivisión. El retículo estaba provisto de una aguja que atravesaba su centro y, por ella, quedaba unido a una varilla que terminaba en una abrazadera que sujetaba estrechamente el tubo del telescopio. Cuando Galileo miraba a través del telescopio con su ojo derecho, veía al mismo tiempo el retículo con el izquierdo y superponía ópticamente Júpiter a la aguja central. Entonces hacía girar el retículo hasta

lograr alinear una línea horizontal con el plano de los satélites. Finalmente, el retículo se acercaba o alejaba del ojo hasta que el disco amplificado de Júpiter encajara exactamente en el espacio comprendido entre dos divisiones verticales, con una tercera división que partiera exactamente el disco por la mitad. Ya que el espacio comprendido entre dos divisiones adyacentes cualesquiera representaba entonces un radio de Júpiter, resultaba fácil medir la posición de un satélite, en unidades iguales a un radio de Júpiter, contando el número de divisiones que le separaban de la aguja. Tales divisiones podían también proporcionar las coordenadas necesarias para medir los ángulos determinados por objetos situados fuera del plano del satélite.

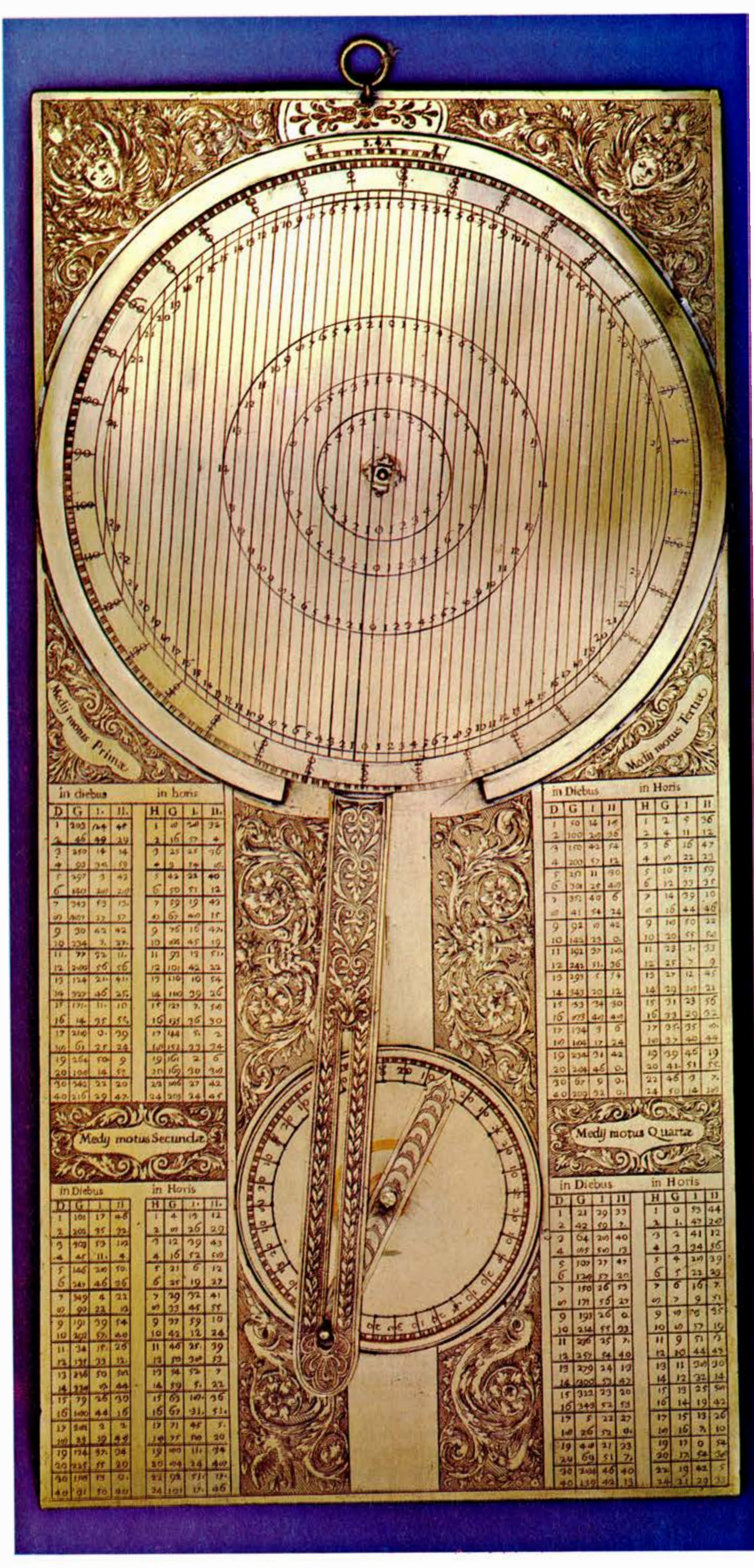
Con el instrumento micrométrico, Galileo podía computar el radio angular aparente del disco de Júpiter utilizando como unidad un segundo de arco. Supongamos que, en una noche determinada, la distancia entre su ojo y el retículo era de un metro. Si asumimos que el espacio entre dos subdivisiones del retículo era de dos milímetros, el radio del disco ampliado de Júpiter sería aquella noche de dos milímetros. Partiendo de estas mediciones y de una tabla de senos, Galileo podía haber calculado que el radio de Júpiter subtendía en el ojo un ángulo de 0,115 grados. Dividiendo 0,115 grados por 18, el número de aumentos del telescopio, se obtiene un radio aparente de 23 segundos de arco.

En este cálculo nos hemos basado en suposiciones relativas al espacio comprendido entre las subdivisiones del retículo y la distancia entre éste y el ojo. En realidad, el diario de Galileo no establece estas medidas, sino la razón que guardan entre sí. Cualesquiera que fuesen sus mediciones, resulta claro que podía dibujar una escala precisa de radios aparentes de Júpiter para cualquier noche. La escala, no obstante, variaría cada noche a medida que se modificara la distancia entre la Tierra y Júpiter. Galileo podía escoger una distancia determinada entre el retículo y el ojo que le pudiera servir de longitud de referencia. En una noche cualquiera podía encontrar la razón existente entre la elongación de un satélite expresada en unidades de retículo, situando éste a la distancia-patrón, y su elongación expresada en radios de Júpiter. La razón proporcionaría una escala para 24 radios de Júpiter (u otra distancia cualquiera) que podía compararse con la escala en otras noches o utilizarse para encontrar la distancia de otros objetos vistos en la misma noche.

En enero de 1612, Galileo encontró que el diámetro aparente de Júpiter era, aproximadamente, de 41,5 segundos de arco. Cuatro meses más tarde, cuando la Tierra y Júpiter se encontraban más alejados, lo midió en 39,25 segundos de arco. Cálculos modernos del diámetro en esos dos días mencionados revelan una razón similar de la reducción visual. Esta similaridad incrementa nuestra confianza en sus mediciones y en nuestra reconstrucción de sus métodos de trabajo.



ESTE *JOVILABIO* DE PAPEL fue diseñado por Galileo para ahorrarse el trabajo de hacer cálculos trigonométricos cuando predecía posiciones de satélites a partir de tablas de sus movimientos. Fue utilizado por Galileo en 1612 y 1613. El círculo exterior, con un radio de nueve centímetros, representa la órbita de Calixto. El círculo está dividido en grados y el diámetro horizontal se divide en 48 partes iguales, cada una de las cuales representa un radio de Júpiter. Las órbitas de los tres satélites interiores corresponden a los tres círculos menores internos. El pequeño círculo situado en el centro es Júpiter. Mediante el jovilabio, la posición tabulada de un satélite, expresada en grados de rotación en torno a Júpiter, puede transformarse en elongación expresada en radios de este planeta. Para realizar este cálculo se pasaba un hilo por el centro del jovilabio y se hacía un nudo por detrás del papel. A continuación se estiraba el hilo formando una línea que correspondiera al grado de rotación previamente calculado. Desde la intersección del hilo con la órbita del satélite se trazaba una perpendicular a la línea horizontal. Entonces podía leerse, sobre la línea, la elongación expresada en radios de Júpiter. Las tablas situadas por encima del jovilabio indican las posiciones de los satélites a intervalos de diez minutos, por cada hora y por cada día. Los valores de las tablas son próximos a los que se aceptan en la actualidad.



JOVILABIO **DE BRONCE que fue construido por Galileo hacia 1617. El jovilabio de bronce se basa en el mismo principio que el instrumento de papel que aparece en la fotografía de la página precedente.**

Aunque la razón medida por Galileo es correcta, los diámetros de Júpiter no lo son, porque su sistema de lentes producía una imagen del planeta con bordes imprecisos. No obstante, podemos deducir el grado de distorsión a partir de los datos que nos ofrece el diario. Galileo anotó que el radio de la órbita del satélite más externo que podía observar era igual a 24 radios de Júpiter; el valor moderno es del orden de 26,4. Su telescopio incrementaba el tamaño aparente de Júpiter 1,1 veces aproximadamente (26,4 dividido por 24). Si aplicamos el coeficiente de corrección 1,1 a las posiciones del satélite, éstas se aproximan a los valores correctos. Este coeficiente de corrección permanece constante sea cual fuere la posición de Júpiter con respecto a la Tierra, ya que el planeta arrastra a los satélites consigo.

Es una suerte para los intentos modernos de reconstrucción que Galileo escogiera un objeto físico –el disco variable de Júpiter– como unidad de medida. Si hubiera adoptado una unidad fija teórica, como el diámetro aparente medio del disco de Júpiter, y lo hubiera utilizado cada noche para registrar la posición de cada satélite, sus mediciones hubieran resultado mucho menos fiables. 40 segundos de arco hubieran constituido un buen valor medio para el diámetro, pero en una noche determinada el diámetro real habría podido variar hasta un 25 por ciento con respecto al mencionado valor medio.

Al basarse en el radio visual real de Júpiter, variable de una noche a otra, Galileo nos ha permitido determinar exactamente lo que medía y compararlo con las observaciones de satélites que puso por escrito. De esta manera hemos determinado el margen típico de error de observación en sus mediciones. Si consideramos el coeficiente de corrección para los límites verdaderos de la imagen de Júpiter, conocemos también el margen de error de observación en las notas de su diario. Utilizando tablas modernas hemos calculado numerosas posiciones de los satélites en las fechas que corresponden a las observaciones de Galileo. Hemos concluido que los errores en las mediciones realizadas en 1612 y 1613 eran frecuentemente inferiores a 10 segundos de arco y sólo rara vez llegaban a 20 segundos de arco.

El telescopio con el que Galileo observó Neptuno tenía, según parece, 18 aumentos y un poder de resolución de 10 segundos de arco. El campo

visual era, probablemente, de 17 minutos de arco, algo más que el radio de la luna llena. El poder de resolución de 10 segundos de arco se ha deducido de la investigación que hemos llevado a cabo acerca de las mediciones de Galileo y está confirmado por dos telescopios y un objetivo, conservados en el Museo de Historia de la Ciencia de Florencia, que posiblemente pertenecieron a Galileo. Este es el poder de resolución máximo y sólo podía alcanzarse al observar puntos brillantes en un campo oscuro. Para observaciones situadas en la zona próxima a Júpiter, el brillo del planeta reducía considerablemente el poder de resolución del instrumento. Rara vez podía Galileo distinguir un satélite hasta que éste se encontraba a una distancia de 25 segundos de arco –más de un radio de Júpiter– con respecto al borde del disco del planeta.

La precisión de una observación astronómica se encuentra limitada no sólo por la exactitud de las mediciones angulares, sino también por la fidelidad con la que se registra el tiempo. Las horas que anota Galileo son, a veces, aproximaciones, aunque una comparación entre sus posiciones de los satélites y los cálculos retrospectivos hechos a partir de tablas modernas muestra que su determinación del tiempo era frecuentemente exacta y resulta fiable, al menos, con un margen de 15 minutos. En la mayoría de los casos, un pequeño error en el tiempo registrado no afectaría de manera significativa a la posición de un satélite. Lo mismo podemos decir acerca de la hora correspondiente a una posición de Neptuno.

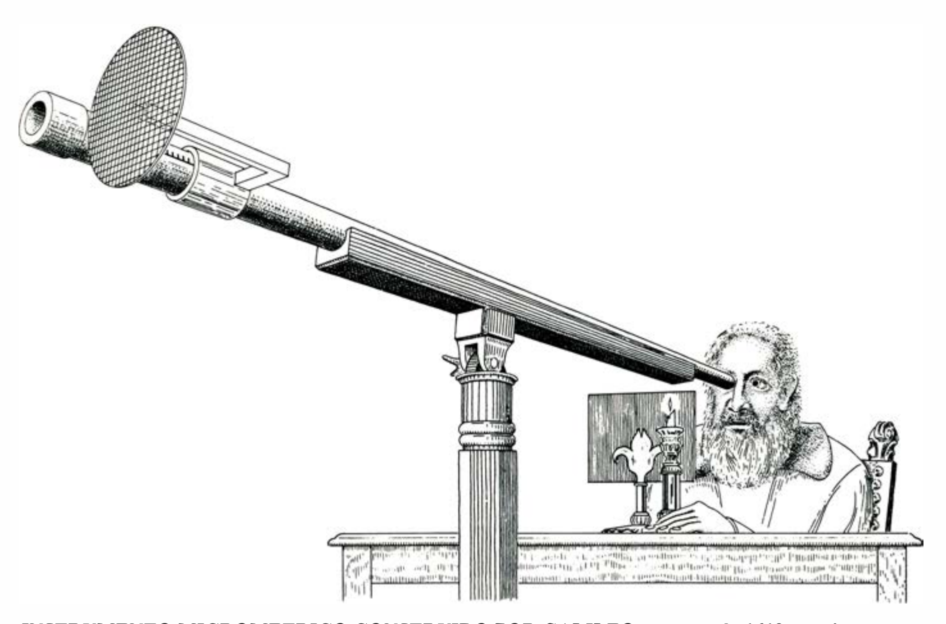
Estos eran los métodos de Galileo para realizar mediciones astronómicas. Hacia las 3,45 de la mañana del 28 de diciembre de 1612 trazó un diagrama de los satélites de Júpiter al que añadió una línea quebrada que formaba un ángulo con el plano de los satélites. Cerca del borde de la página dibujó una estrella en el extremo de la línea quebrada y anotó sobre ella *fixa* (el término latino que designa a las estrellas fijas). El objeto así señalado, situado al sur y al este de Júpiter y al norte del plano muy inclinado de los satélites, no era una estrella fija sino el planeta Neptuno. En aquel momento se encontraba prácticamente estacionario con respecto al fondo de estrellas porque su movimiento aparente, visto desde la Tierra, estaba cambiando de sentido, de directo a retrógrado. Aquella misma madrugada, un poco más tarde, Galileo hizo una nueva observación de los satélites y volvió a anotar la presencia del mismo objeto.

¿Cómo podemos tener la seguridad de que el cuerpo celeste al que Galileo denominó *fixa* era Neptuno? En primer lugar, no hay duda de que pudo ver el planeta si su telescopio apuntaba en la dirección conveniente. Neptuno tiene una magnitud visual de 7,8 y Galileo había discernido las posiciones de estrellas de brillo mucho más débil. Por otra parte, tampoco hay duda de que Neptuno se encontraba dentro de su campo visual. Formaba el mismo ángulo con el plano de los satélites que la estrella fija que dibujó Galileo. En tercer lugar, tiene especial importancia el constatar que no pudo observar, con su telescopio, ninguna estrella fija que se encontrara sobre la línea que partía del plano de los satélites.

Aquella noche, por vez primera, Galileo dibujó una línea quebrada en las notas de su diario. Probablemente la dibujó quebrada para indicar que no representaba distancia sino dirección. Neptuno estaba entonces a 41 radios jovianos al este de Júpiter y si lo hubiera representado utilizando la misma escala que para los satélites, se hubiera encontrado fuera de los límites de la estrecha página.

En los centenares de observaciones registradas que Galileo llevó a cabo antes del 28 de diciembre de 1612, sólo encontramos un escaso número de casos en los que añadió una estrella fija a su diagrama de las lunas de Júpiter. La razón por la que anotó estas estrellas y no otras resulta evidente. En cada ocasión tomó nota de la estrella por primera vez cuando Júpiter se aproximaba a ella y estaba a punto de adelantarla. Al seguir el movimiento de Júpiter durante las pocas noches en las que una estrella permanecía en el mismo campo visual que el planeta, Galileo utilizaba un método independiente para comprobar sus mediciones del diámetro aparente de Júpiter. Las estrellas servían de puntos de referencia que le ayudaban a investigar no sólo los movimientos relativos de Júpiter y sus lunas sino también sus movimientos absolutos. Existían, por tanto, precedentes y razones suficientes para incluir un objeto como Neptuno en sus diagramas de las lunas de Júpiter.

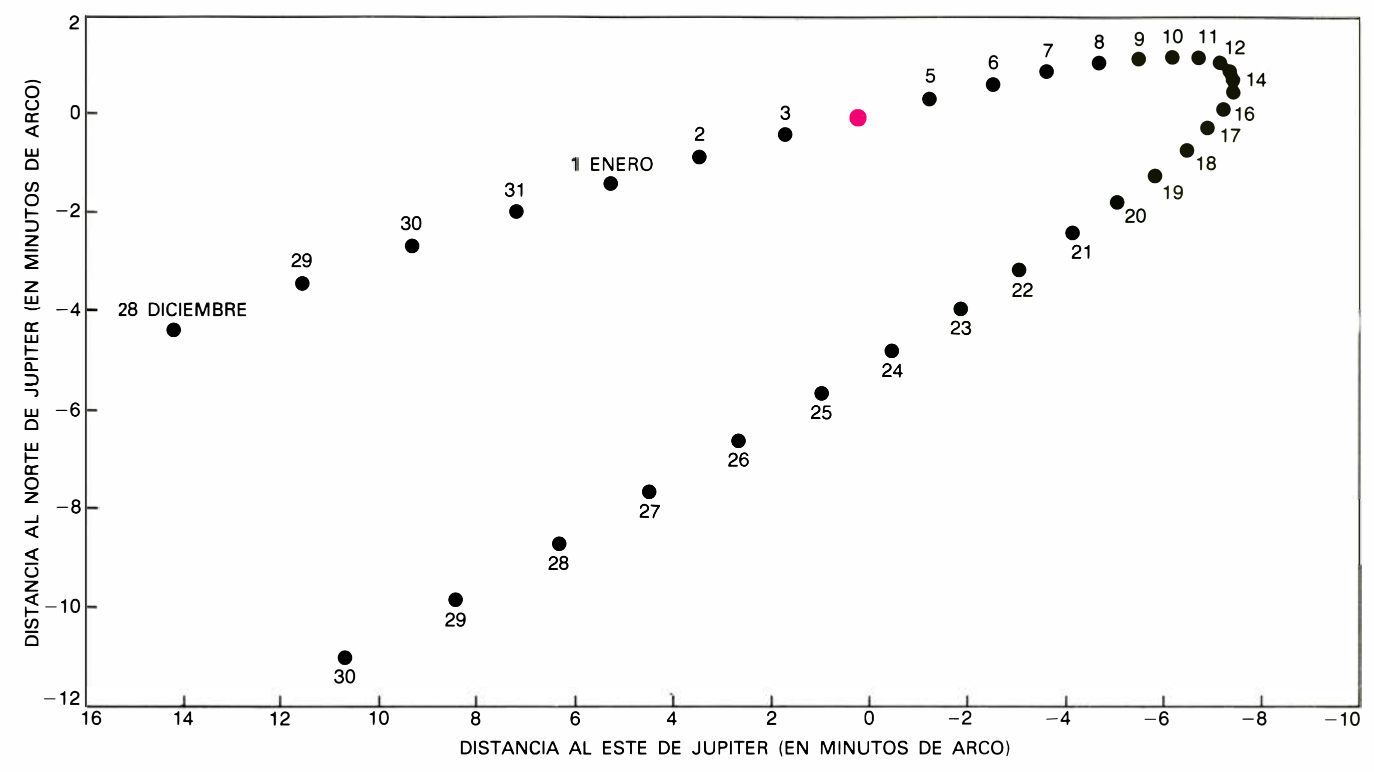
La siguiente observación de Galileo, no obstante, incluía algo sin precedentes y que nunca volvió a aparecer en los registros de sus observaciones. Entre el 28 de diciembre y el 1 de enero el cielo estaba demasiado nublado para poder llevar a cabo observaciones fia-



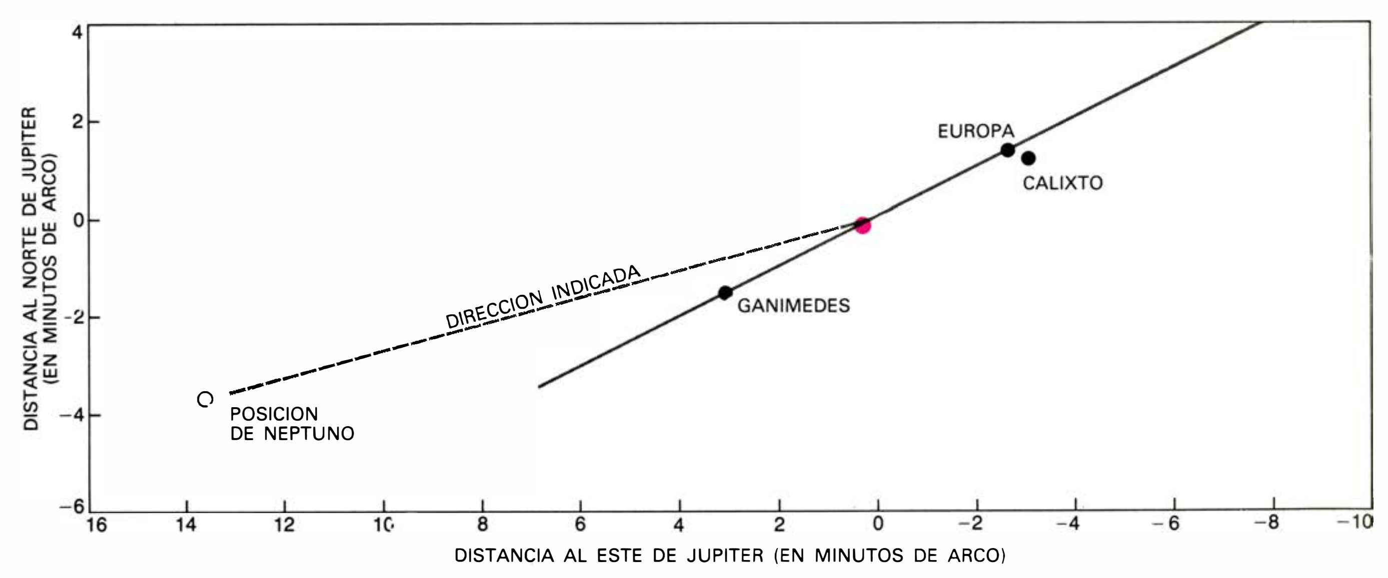
INSTRUMENTO MICROMETRICO CONSTRUIDO POR GALILEO en enero de 1612, que incrementaba la precisión de sus observaciones telescópicas. El micrómetro, que no se conserva, fue rara vez mencionado por Galileo pero probablemente funcionaba como el que aparece en la figura. El micrómetro era un retículo circular montado cerca del extremo del telescopio. El retículo estaba formado por líneas separadas por espacios iguales. Galileo miraba a través del telescopio con su ojo derecho y al retículo con el izquierdo, con lo cual superponía ópticamente Júpiter y sus satélites sobre el retículo. A continuación, hacía girar el retículo hasta conseguir que una línea horizontal se alineara con el plano de los satélites de Júpiter, y lo aproximaba o lo alejaba del ojo hasta que el planeta quedaba perfectamente encajado entre dos líneas verticales mientras una tercera vertical lo dividía por la mitad. Ya que, por consiguiente, el espacio comprendido entre dos líneas adyacentes cualesquiera equivalía a un radio de Júpiter, la elongación de un satélite, expresada en radios del planeta, podía encontrarse contando las líneas divisorias existentes comprendidas entre el satélite y Júpiter. La distancia entre el retículo y el ojo podía medirse mediante un listón graduado situado sobre un lado del telescopio. El retículo estaba unido al telescopio por una abrazadera que podía moverse a lo largo del listón graduado. El espacio que estaba comprendido entre dos divisiones del listón era el doble del espacio entre dos divisiones del retículo circular, de tal modo que el mencionado espacio era equivalente al diámetro aparente de Júpiter.

bles. El 2 de enero el cielo se había despejado e hizo su observación singular a medianoche. Registró una estrella fija al oeste y al norte de Júpiter que carecía de utilidad para su estudio del movimiento del planeta ya que éste la había ya adelantado. La estrella se encontraba tan alejada de Júpiter que difícilmente pudo lograr situar a ambos cuerpos celestes en el mismo campo visual y, de hecho, cinco horas más tarde no lo pudo conseguir. Por consiguiente, debió haber realizado un esfuerzo especial que motivó el que descubriera la posición de esta estrella.

No tenemos ninguna otra evidencia de que Galileo buscara, alguna vez, otra estrella fija junto a la cual ya hubiera pasado Júpiter. ¿Por qué actuó así en esta ocasión? Podemos afrontar la cuestión si consideramos en qué lugar esperaba ver la estrella fija del 28 de diciembre (o sea Neptuno) cuando de nuevo la buscó cinco días más tarde. En su primera observación encontró a Neptuno al este de Júpiter, a una distancia igual al doble de la elongación máxima de Calixto. Ya que una estrella fija necesita, habitualmente, cuatro o cinco días para recorrer esta distancia en relación a Júpiter, Galileo probablemente esperaba encontrarla muy cerca de este planeta el dos de enero. De hecho estaba allí pero, debido a alguna razón, Galileo parece no haberla visto. Pudo, tal vez,



POSICIONES DE NEPTUNO calculadas retrospectivamente entre el 28 de diciembre de 1612 y el 30 de enero de 1613. Los puntos negros corresponden a Neptuno y el de color a Júpiter. El 4 de enero, Júpiter ocultó a Neptuno. La siguiente ocultación similar no se produjo hasta el mes de septiembre de 1702.



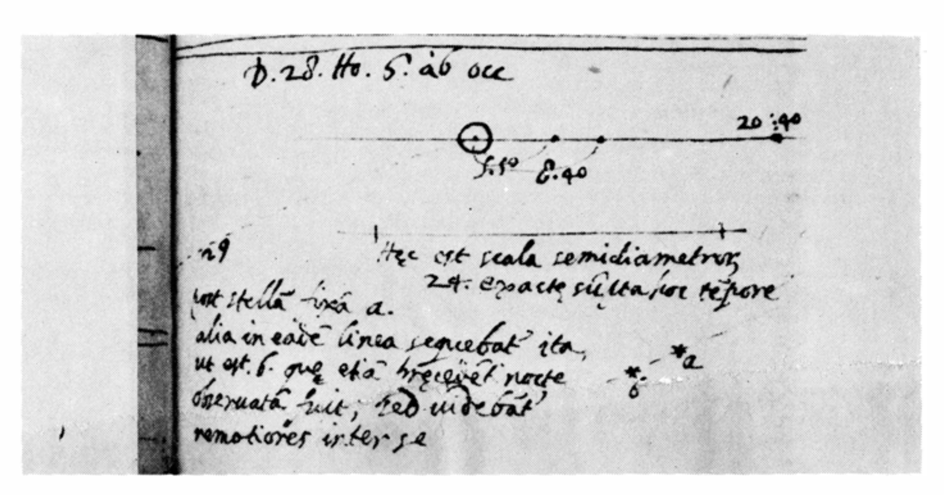
PRIMERA OBSERVACION DE NEPTUNO, ocurrida el 28 de diciembre del año 1612. Galileo trazó una línea quebrada que representaba la dirección del cuerpo celeste que identificaba con una estrella fija pero que, en realidad, era Neptuno. No señaló la posición, sino la dirección en la que se encontraba porque el planeta hubiera debido representarse fuera de la estrecha página de su libro de notas, si hubiera intentado situarlo de acuerdo con la misma escala que utilizaba para situar las posiciones de los satélites. A partir de los datos de Galileo, los autores han determinado dónde debió ver Neptuno aquella noche: unos cuatro minutos de arco al norte y trece minutos al este de Júpiter. En diagramas anteriores no solía añadir estrellas fijas a las lunas jovianas.

observar una estrella fija, situada muy hacia el oeste, al recorrer el cielo en busca de Neptuno. Esta estrella era más brillante que el planeta y se encontraba al norte del plano de los satélites. En cualquier caso, la estrella fija que vio aquella noche (la que ahora denominamos SAO 119234) acabaría por desempeñar un papel muy principal en la siguiente observación de Neptuno registrada.

Hemos discutido mucho la causa por la que Galileo no pudo descubrir la posición de Neptuno el 2 de enero. Tal vez la cobertura de nubes era de tal índole que pudo observar SAO 119234 pero no Neptuno, de brillo mucho más débil: de una nota tachada correspondiente a una noche de diciembre de 1612, en la que aparece la frase "dudoso debido a las nubes", se desprende que intentó realizar observaciones de satélites incluso en condiciones atmosféricas poco favorables. Otra posibilidad es que Neptuno se encontrara demasiado cerca de Júpiter o de una de sus lunas para poder ser distinguido. No existe prueba alguna de esta posibilidad; de hecho, tanto la órbita calculada como la siguiente observación de Neptuno realizada por Galileo nos llevan a la conclusión de que el astrónomo pudo ver el planeta el 2 de enero. No obstante, ya que ponemos en cuestión la órbita calculada de Neptuno debido a la observación de Galileo, no podemos descartar la posibilidad de que este planeta se hubiera visto ocultado por uno de los satélites del sistema de Júpiter, por más que tal hipótesis sea poco probable.

Entre el 6 y 19 de enero el tiempo, una vez más, parece haber impedido a Galileo realizar nuevas observaciones. No pudo ver el cambio de sentido en el movimiento de Júpiter, de este a oeste, que tuvo lugar el 14 de enero. Una semana más tarde, este mismo planeta, que se movía hacia el norte así como hacia el oeste, volvió a alcanzar a Neptuno y a SAO 119234, los cuales se encontraron, en lo sucesivo, al sur del plano de los satélites.

El 20 de enero Galileo reemprendió sus observaciones y el 25 de enero registró de nuevo la posición de SAO 119234. En primer lugar anotó que lo había visto a diez radios jovianos por debajo de Júpiter, sobre una línea perpendicular al plano de los satélites. Ya que tanto las líneas como los ángulos rectos resultan fáciles de determinar, Galileo anotó con frecuencia alineaciones perpendiculares. La posición registrada está de acuerdo con la que se obtiene mediante el cálculo



APUNTACION DEL LIBRO DE NOTAS correspondiente al 28 de enero de 1613, reproducido a escala real. De acuerdo con lo escrito, Galileo había detectado el cambio, que se había producido en una sola noche, en las posiciones relativas de Neptuno y una estrella fija (que ahora se designa como SAO 119234). La nota escrita abajo a la izquierda dice: "Más allá de la estrella fija *a* le seguía otra, en la misma línea, tal como [hace] *b*, que fue observada también la noche anterior, aunque [entonces] parecían encontrarse más separadas". La página del libro de notas fue proporcionada por la Biblioteca Central Nacional.

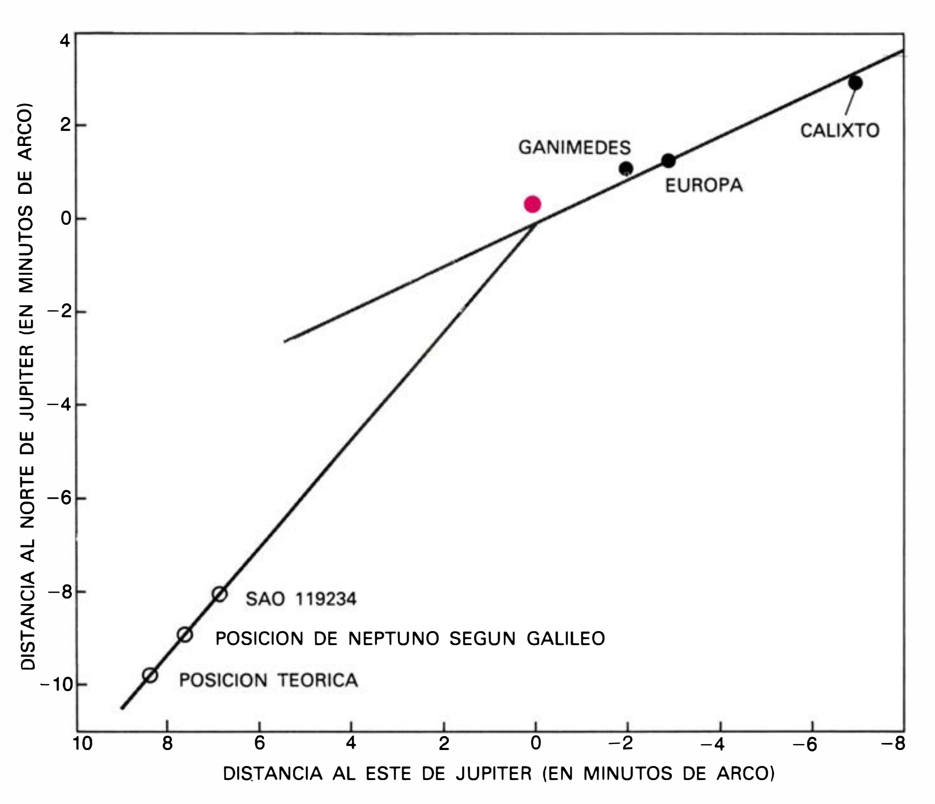
retrospectivo. Galileo podía también haber situado en el mismo campo visual a Neptuno, junto con Júpiter y SAO 119234, pero el planeta era poco brillante, estaba alejado de Júpiter y no se encontraba en una alineación perpendicular o en línea recta.

Entre el 25 y el 27 de enero Galileo anotó cada noche la posición de SAO 119234. El 28 de ese mes lo situó a una distancia de 29 radios jovianos (o 32 radios si aplicamos el coeficiente de corrección) desde el centro de Júpiter. Esta posición de la estrella, tanto en distancia como en dirección con respecto a Júpiter, concuerda de manera muy aproximada con los cálculos basados en datos modernos.

Aquella noche, a las 11, Galileo situó por última vez la posición de Neptuno. La anotó porque se encontraba más allá de SAO 119234 sobre una línea recta que pasaba por la estrella y por el centro de Júpiter. Había observado Neptuno también la noche anterior, pero no había registrado su posición. El día 28 comentó que, en la noche de la víspera, Neptuno y SAO 119234 parecían encontrarse más separados. Ya que las posiciones relativas de las estrellas fijas no pueden cambiar de la noche a la mañana, debió sentirse perplejo al recordarlo. Tenía buenas razones para confiar en su memoria. Durante tres años se había dedicado a señalar las posiciones relativas de los satélites de Júpiter, razón por la cual se había sensibilizado a los cambios bruscos de posición. Nuestros cálculos muestran que, del 27 al 28 de enero, Neptuno había avanzado 2,5 radios de Júpiter en dirección hacia SAO 119234. El día 27 la separación entre el planeta y la estrella era de 6,25 radios de Júpiter, por lo cual esta aproximación de 2,5 radios en una noche implicaba que la separación entre ambos cuerpos se había reducido en un 40 por ciento.

El día 28, Galileo dibujó con detalle las posiciones de Júpiter y de sus satélites. Incluyó a SAO 119234 en el diagrama, trazó una línea que lo conectaba con el centro de Júpiter e indicó la distancia. Entonces, por vez primera en su diario, dibujó una "escala exacta" de 24 radios de Júpiter. La escala de 24 radios de este planeta mide 47 milímetros, lo que apoya nuestra reconstrucción de su retículo, cuyas líneas mantendrían entre sí una separación de 2 milímetros. Galileo no intentó situar a Neptuno en el diagrama porque no hubiera cabido, a esta escala, en la página. En lugar de ello, realizó un dibujo independiente y paralelo de SAO 119234 y de Neptuno, trazado a la misma escala, en el que ambos cuerpos mantienen una distancia de 3,75 radios. Añadió una nota a este dibujo en el que resumía su observación: "Más allá de la estrella fija *a* le seguía otra en la misma línea, tal como [hace] *b*, que fue observada también la noche anterior, aunque [entonces] parecían encontrarse más separadas".

El trabajo que se tomó Galileo para realizar un diagrama a escala de Neptuno y SAO 119234 ha sugerido a uno de los autores de este artículo (Drake) que el astrónomo italiano pretendía confirmar su primera impresión de movimiento aparente relativo mediante observaciones en las noches siguientes. Pero no las llevó a cabo. Pueden expresarse serias dudas acerca de la hipótesis, antes expuesta, de que pre-



SEGUNDA OBSERVACION DE NEPTUNO (la del 28 de enero de 1613), que muestra la discrepancia entre la observación de Galileo y la órbita calculada hoy. Hay pocos motivos para dudar de la exactitud de sus observaciones y, por otra parte, tenemos razones para creer que la órbita puede ser errónea. El descubrimiento de Neptuno en el siglo XIX demostró la validez de la mecánica celeste newtoniana.

tendía proseguir este tipo de estudio, dada la ausencia de todo registro de la posición de SAO 119234 durante la noche siguiente, a pesar de que los cálculos modernos muestran que podía haber abarcado en su campo visual a Neptuno, SAO 119234 y Júpiter. Es obvio que no podemos tener absoluta confianza en la posición de Neptuno basada en cálculos modernos mientras subsistan incertidumbres acerca de su órbita exacta. Tal vez el mal tiempo obstaculizó de nuevo su trabajo. Galileo observó satélites aquella noche, pero éstos son mucho más brillantes que Neptuno y SAO 119234, los cuales podían haberse visto ocultados por un cielo nuboso. De cualquier modo, la noche del 28 de enero fue su última oportunidad para ver estos cuerpos celestes ya que, en lo sucesivo, no podría ya situar de nuevo a los tres objetos en el mismo estrecho campo visual. Sin un telescopio provisto de un círculo graduado en su montura y sin poder contar con Júpiter como punto de referencia, difícilmente podía encontrar Galileo un planeta y un satélite que no son discernibles a simple vista y, obviamente, no tenía medios para determinar sus movimientos.

A no ser por el método pionero utilizado por Galileo para convertir su telescopio en un instrumento de medición, sus observaciones de Neptuno no habrían tenido repercusión ninguna en la astronomía moderna. No obstante, al registrar las distancias en radios de Júpiter y dibujar diagramas precisos, nos ha proporcionado una posición de Neptuno en el siglo XVII que es digna de confianza.

En el diagrama de Galileo, la distancia entre Neptuno y SAO 119234 es poco más de un minuto de arco. La órbita de Neptuno que se acepta habitualmente lo sitúa a unos dos minutos de la estrella. La posible discrepancia, de casi un minuto de arco, entre la posición registrada por Galileo y la calculada retrospectivamente constituye una diferencia notable para las pautas de precisión de la mecánica celeste. Tras realizar todo tipo de ajustes razonables y dejar un margen al error de observación, la discrepancia es lo suficientemente grande como para dudar acerca de la exactitud de la órbita aceptada. Todo ello sugiere que alguna masa aún por identificar puede estar perturbando el movimiento de Neptuno. Una posibilidad puede estar constituida por la existencia de un planeta no descubierto todavía. El nuevo planeta debería encontrarse más allá de Neptuno ya que, de otro modo, afectaría de modo significativo a las órbitas de los planetas interiores.

Es curioso observar que el descubrimiento de Neptuno en el siglo XIX se debió, en sí mismo, a una irregularidad orbital. Urano, descubierto en 1781, había mostrado, durante el medio siglo siguiente, discrepancias, con respecto a las posiciones de su órbita, que no podían justificarse en razón de las fuerzas gravitatorias de los planetas conocidos. Se sugirió entonces que algún planeta, aún por descubrir, podía estar atrayendo a Urano, de modo que afectara en grado menor a las órbitas de los primeros seis planetas. John Couch Adams, en Inglaterra, y Urbain Jean Joseph Leverrier, en Francia, investigaron la posibilidad de que existiera un nuevo planeta. Con puntos de partida sorprendentemente idénticos, sus investigaciones les llevaron a conclusiones muy semejantes en la práctica, que se referían a la masa probable del planeta desconocido, a su distancia del Sol y a la región del cielo en la que los astrónomos debían buscarlo.

Adams envió su trabajo a George B. Airy, Astrónomo Real de Inglaterra, pero éste lo dejó de lado y se abstuvo de realizar la labor que le había sido encomendada. Mientras tanto, Leverrier había obtenido la cooperación del Observatorio de Berlín en el que se acababa de realizar una nueva serie de mapas estelares. Galle, que trabajaba allí, buscó, en la región del cielo que Leverrier había sugerido, algún objeto que no figurara en los mapas. Lo encontró, en la primera noche de observaciones, y a la siguiente, descubrió que se había desplazado un poco. El objeto era, desde luego, Neptuno.

Las noticias acerca del descubrimiento de Neptuno se difundieron rápidamente y causaron sensación a nivel internacional: algo similar había sucedido cuando Galileo anunció, por primera vez, en 1610 que había visto nuevos "planetas" a través de su telescopio. En su tiempo todas las estrellas errantes recibían el nombre de planetas y las descubiertas por Galileo eran los satélites de Júpiter. El carácter inesperado de la aparición de tales satélites hizo que su descubrimiento constituyera una noticia sensacional. La sensación motivada por la aparición de Neptuno tuvo su origen en una razón de signo contrario: se esperaba que existiera un planeta como consecuencia de las leyes de Newton sobre el movimiento y la gravitación y debido al carácter muy preciso de las mediciones realizadas. El descubrimiento constituyó, por tanto, una demostración espectacular del éxito de la mecánica celeste. La tarea que queda ahora es comprobar las cuidadosas mediciones llevadas a cabo por Galileo, volviendo a calcular la órbita de Neptuno.

Asimetría cósmica entre materia y antimateria

Según parece, el universo actual está formado, en casi su totalidad, por materia. Pruebas de orden cosmológico y de la física de partículas nos permiten esbozar ya una explicación

Frank Wilczek

Todos los constituyentes fundamentales de la materia están apareados: para cada clase de partícula existe una antipartícula, de masa idéntica, pero con otras propiedades opuestas como, por ejemplo, la carga eléctrica. La unión de las dos grandes teorías físicas del siglo XX, a saber, la relatividad y la mecánica cuántica, exige el apareamiento simétrico de partículas y antipartículas. La simetría ha superado con éxito la prueba de la experimentación. Desde 1932, cuando se descubrió el positrón (o antielectrón), el catálogo de las antipartículas ha crecido al mismo ritmo que el nomenclátor de las partículas ordinarias. De hecho, se ha asistido con frecuencia a la detección simultánea de una partícula y su antipartícula al crearse las dos, como un par, en una colisión de alta energía en un acelerador de partículas. Tales colisiones parecen producir siempre materia y antimateria en cantidades iguales. Y así, durante mucho tiempo se supuso que las leyes de la naturaleza no expresaban preferencia alguna por la materia o por la antimateria.

A pesar de todo lo cual, no se encuentra antimateria casi nunca fuera del laboratorio. Los átomos que componen la tierra están formados por neutrones, protones y electrones, pero nunca de sus antipartículas. ¿Permanece esta asimetría en todos los puntos del universo? Es decir, ¿consta el universo predominantemente de materia, con muy poca antimateria? Si así es, ¿existió siempre la asimetría, o empezó el universo con igual número de partículas y antipartículas y más tarde se desarrolló de alguna manera un desequilibrio?

Hallazgos recientes en los campos de la cosmología y de la física de partículas aportan respuestas a esas cuestiones. Sugieren que, en el primer instante después de la gran explosión ("big bang"), cuando el universo era mucho más caliente y denso que ahora, había cantidades iguales de materia y antimateria. Pero antes de que hubieran transcurrido los primeros 10^{-35} segundos, las violentas colisiones entre partículas crearon condiciones que condujeron prontamente a una asimetría entre materia y antimateria. Desde entonces, la asimetría se ha asentado en el universo. Se ha llegado a semejante conclusión a través de un camino cuyo firme se resiente en varios puntos, pero intentaré demostrar que es la vía correcta.

¿Cómo podemos estar seguros de que el universo está enteramente formado por materia? Es fácil demostrar que la materia y la antimateria no pueden mezclarse homogéneamente. Siempre que una partícula y la correspondiente antipartícula se juntan, se aniquilan entre sí y su masa se convierte en energía. Por tanto, una estrella formada mitad por materia y mitad por antimateria desaparecería inmediatamente en una explosión titánica. Existe, sin embargo, la posibilidad de que la materia y la antimateria coexistan en el universo si cada una está confinada en regiones aisladas con espacio vacío por medio

Una indicación de la preponderancia de materia sobre antimateria viene proporcionada por los rayos cósmicos, las partículas de alta energía que llegan del espacio. Invariablemente, parecen ser partículas de materia, tales como protones y electrones y núcleos atómicos formados por protones y neutrones; no se observan antipartículas. Aunque no se conoce en todos sus detalles el origen de los rayos cósmicos, se sabe con certeza que proceden de toda la galaxia y algunos de ellos pueden tener incluso una fuente más lejana. Por tanto, es coherente pensar que la Vía Láctea consta, en su integridad, de materia y también es razonable, aunque con menos solidez, deducir que el grupo de galaxias al que pertenece la Vía Láctea está formado a su vez, y en toda su extensión, por materia.

Ya resulta más difícil asegurar que galaxias más lejanas estén formadas por materia. El mirar simplemente a una galaxia no ofrece ninguna indicación sobre si en su composición entra materia o antimateria. "Mirar" una galaxia implica la detección de fotones, o cuantos de radiación electromagnética. Los fotones comprenden, además de los de la luz visible, los de las ondas de radio, rayos X, rayos gamma, etcétera. El problema reside en que el fotón es su propia antipartícula, y no hay manera de distinguir un fotón emitido por materia de otro emitido por antimateria. En consecuencia, la luz de una galaxia de antimateria sería idéntica a la de una galaxia de materia, incluso en la estructura detallada del espectro. Por ejemplo, las líneas de emisión características del átomo de hidrógeno se verían duplicadas con exactitud por las líneas de emisión características del átomo de antihidrógeno.

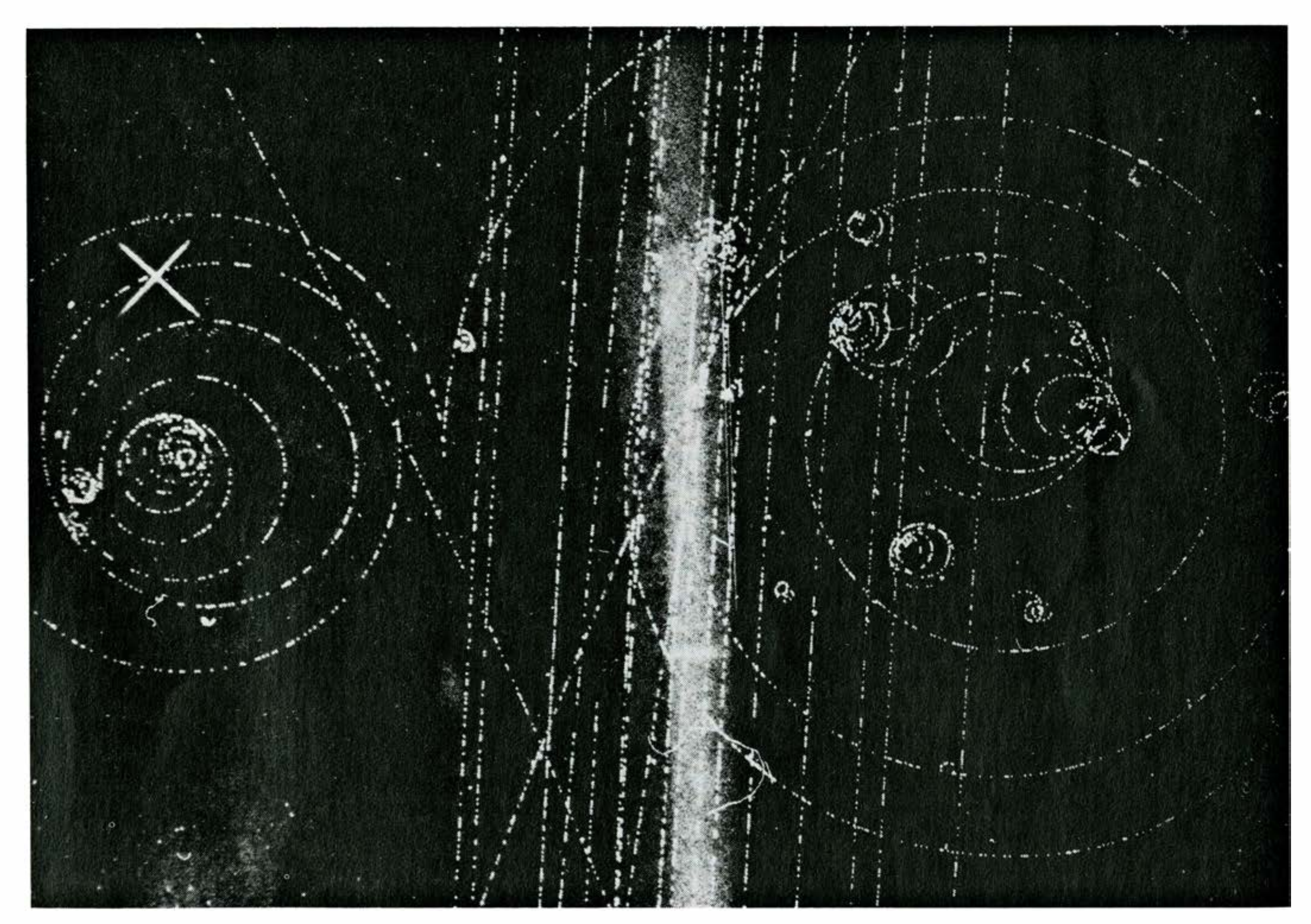
Hay un caso en que las observaciones de fotones podrían revelar indirectamente la presencia de antimateria. Si una galaxia de antimateria estuviera cercana a una galaxia de materia, la región límite entre ellas sería escenario de frecuentes aniquilaciones partícula-antipartícula. La energía de cada una de tales aniquilaciones aparecería, con el tiempo, en forma de fotones en las longitudes de onda de los rayos gamma. La región fronteriza sería así un lugar donde se emitiría copiosamente radiación gamma. Se conocen fuentes astronómicas de radiación gamma y se están estudiando, pero no se ha encon-

trado ninguna fuente con las características apropiadas. Sin embargo, este argumento pierde validez si el espacio que separa la materia de la antimateria está vacío. En el mejor de los casos, la no observación de fuerte emisión gamma sugiere que los agregados de galaxias deben estar formados enteramente por materia o enteramente por antimateria, pero no por una mezcla de las dos. Los agregados están inmersos en gas intergaláctico y cualquier diferencia de composición dentro de un agregado daría lugar a radiación gamma.

En el futuro, la pregunta de si existen en el universo agregaciones sustanciales de antimateria podrá responderse con la fabricación de telescopios que no detecten fotones, sino neutrinos. Al revés del fotón, el neutrino tiene una antipartícula distinguible. Los neutrinos y los antineutrinos se emitirían en proporciones diferentes en las reacciones nucleares de materia o antimateria. Una estrella compuesta por materia radía principalmente neutrinos, mientras que una estrella compuesta de antimateria daría lugar principalmente a antineutrinos. La incógnita no ha sido despejada aún mediante observaciones de neutrinos, habida cuenta de que construir un telescopio de neutrinos constituye un proyecto de formidable complejidad. Los neutrinos, de masa despreciable, apenas si llegan a interaccionar en absoluto con la materia. Su detección resulta problemática.

Por ahora al menos, la opinión mayoritaria entre astrónomos y astrofísicos es que la materia domina sobre la antimateria en el universo actual. Como he insinuado, las pruebas a favor de este punto de vista no son concluyentes, aunque hay una notable falta de datos que apoyen la existencia de antimateria. Lo que, en definitiva, parece decisivo es la dificultad de imaginar cómo la materia y la antimateria del universo primitivo pudieron haberse segregado en regiones distintas. Parecería más probable que se hubiesen aniquilado simplemente entre sí y en todas partes.

Si el universo está formado, hoy, en su casi totalidad, por materia, uno se siente inclinado a preguntar cómo surgió tal asimetría. Una primera posibilidad sería que la preferencia por la materia estuviera ya en el principio, que el material primordial resultante de la gran explosión fuese predominantemente materia. Aunque esta hipótesis no puede descartarse, al menos de momento, no resulta muy satisfactoria. En principio, cualquier composición del universo podría explicarse de la misma manera. Además, la hipótesis de un desequilibrio primordial concede un papel fundamental a un conjunto de condiciones iniciales que no parecen verosímiles; cualquier número de alternativas parece igualmente plausible. Si pudiéramos construir una teoría coherente con los principios físicos establecidos, que implicara un universo inicial simétrico, resultaría mucho más atrac-

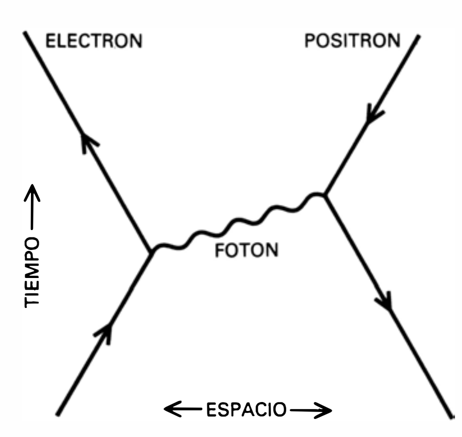


CREACION DE ANTIMATERIA. Se hace visible en una cámara en la que la trayectoria de cualquier partícula que tenga una carga eléctrica se marca por un trazo de burbujas en helio líquido. Aquí, la antimateria es un positrón, cuya trayectoria en espiral en el sentido de las agujas del reloj llena los dos tercios de la derecha de la fotografía. La espiral más pequeña en el sentido contrario a las agujas del reloj representa la trayectoria de un electrón. El positrón es la antipartícula del electrón: aunque idénticas ambas por razón de su masa, en varios otros aspectos (la carga eléctrica, por ejemplo) sus atributos son opuestos. El positrón y el electrón se crearon como un par en la desintegración de un fotón, o cuanto de radiación electromagnética. No puede verse la trayectoria del fotón porque los fotones carecen de carga eléctrica y no producen burbujas en el helio. Se aplicó un campo magnético a la cámara para curvar las trayectorias de las partículas. En los experimentos de alta energía, la creación de pares partícula-antipartícula es corriente, aunque el universo macroscópico parece estar formado predominantemente de materia. Fotografía realizada por Nicholas P. Samios, del Brookhaven Laboratory.

tiva. Precisamente una teoría tal nos viene ofrecida por la conjunción de la cosmología y la física de partículas.

Acontecimiento crucial en la moderna cosmología fue el descubrimiento, en la década de 1920 y por Edwin P. Hubble, de que las galaxias distantes se están alejando de la tierra con velocidades proporcionales a sus distancias. La recesión de las galaxias implica que todo el universo está en expansión. La extrapolación hacia atrás en el tiempo conduce a la conclusión de que, hace unos 10.000 millones de años, el material que ahora forma las galaxias surgió de una explosión de un estado altamente comprimido. Más. El seguimiento de la evolución hacia atrás hasta su límite matemático sugiere que todo el universo fue, en su origen, un punto carente de dimensiones.

En el instante de la gran explosión la densidad y la temperatura del universo eran infinitas. Aunque la temperatura bajó rápidamente, durante el primer minuto superó los 10^{10} grados Kelvin. En ese marco, todo átomo que pudiera haberse formado se hubiese desintegrado en seguida; ni siquiera los núcleos atómicos hubiesen podido sobrevivir: se hubiesen descompuesto en sus partículas constituyentes. En otras palabras, el universo en sus primeros momentos fue un plasma caliente constituido por partículas libres, muchas de las cuales, como los electrones y los protones, portaban carga eléctrica. Puesto que las partículas cargadas en movimiento emiten radiación electromagnética, el universo primitivo hubo de ser rico en fotones.



FUERZA ELECTROMAGNETICA. Puede representarse actuando a través del intercambio de un fotón entre dos partículas cargadas eléctricamente. La interacción se esquematiza en el gráfico mediante un diagrama de Feynman: el tiempo avanza de abajo arriba y el espacio, unidimensional, corre de izquierda a derecha. La línea inferior izquierda representa un electrón. La línea inferior derecha representa un positrón. Las flechas señalan la dirección en que fluye la carga. (En la descripción matemática de este tipo de interacción, un positrón que se desplaza hacia adelante en el tiempo equivale a su antipartícula, el electrón, moviéndose hacia atrás en el tiempo.) El fotón intercambiado altera la trayectoria de cada partícula.

El universo en expansión se enfrió de manera muy semejante a como se enfría un gas en expansión; al cabo de unos tres minutos de la gran explosión, los protones y neutrones empezaron a combinarse para formar los núcleos de átomos de helio. Los restantes protones no combinados acabarían por convertirse en núcleos de hidrógeno. (Todos los elementos más pesados, que son bastante raros en una escala cósmica, se han formado a partir del hidrógeno y el helio en los centros de las estrellas y en las explosiones de supernovas.) Basados en las hipótesis más sencillas que pueden formularse acerca de las condiciones en el universo primitivo y que sean concordes con las leyes físicas conocidas, cabe deducir que la razón de helio a hidrógeno fue, aproximadamente, de uno a tres en peso. Dicha cifra se ajusta a la razón estimada para el universo actual. El éxito de esta predicción atestigua nuestra comprensión de cómo era el universo, minutos después de su nacimiento.

Transcurridos 10.000 años de expansión, el universo era lo suficientemente frío para que la última de las partículas libres cargadas fuera incorporada en los átomos. Cada átomo es eléctricamente neutro porque tiene igual número de cargas positivas y negativas. Los fotones interaccionan sólo débilmente con la materia neutra; a partir de entonces, pues, la materia y la radiación electromagnética en el universo quedaron esencialmente desacopladas. Desde ese momento, la radiación ha secundado de un modo libre la expansión del universo, enfriándose constantemente. En primer lugar, ¿cómo puede enfriarse la radiación? ¿cómo puede tener siquiera una temperatura? Si la radiación se contempla como un gas de fotones, se enfría por expansión, de manera parecida a un gas de partículas materiales, a medida que la energía media de los fotones decrece. Si la radiación se contempla como una onda, la expansión del espacio produce un aumento en la distancia entre dos crestas de onda inmediatas. La longitud de onda mayor corresponde a una menor energía del fotón.

En 1964 se descubrió que la radiación de microondas está incidiendo en nuestro planeta desde todas las direcciones. La radiación corresponde a un gas de fotones que llena el universo con una densidad de unos 300 fotones por centímetro cúbico. La temperatura de la radiación es de 2,7 grados K, un valor mucho menor que la temperatura de 10.000 grados existente en la época del desacoplo. La presencia de la radiación aumenta la base argumental que defiende que esta reconstrucción teórica del universo primitivo es la correcta. Afianzados en esos éxitos, vale la pena intentar extrapolar hacia atrás hasta los primeros instantes del universo, a fin de ver si las condiciones extremas que prevalecían entonces pueden explicar la presente asimetría entre materia y antimateria.

En los primeros segundos, muy pocos, del universo, las partículas del caliente gas primigenio tenían una energía media que superaba las posibilidades de hasta los mayores aceleradores de partículas modernos. Las interacciones de partículas a aquellas energías pudieron haber sido cualitativamente distintas de cuantas podemos observar ahora. Aun cuando los sucesos en el universo primitivo difiriesen en carácter de los accesibles actualmente, cabe suponer, sin embargo, que las leyes de la naturaleza que gobiernan los sucesos han permanecido invariantes. Se precisa, pues, una teoría que prediga cómo actúan las partículas a muy alta energía, basada en las leyes naturales deducidas de los sucesos a energías mucho más bajas.

Entre las leyes naturales a que aludimos están las leyes de conservación aplicadas a los números cuánticos. Un número cuántico es un registro idóneo, que nos sirve de recurso para seguir las distintas propiedades de las partículas. Así, la carga eléctrica puede expresarse como un número cuántico; al protón se le asigna un valor de + 1, al electrón un valor de − 1 y al fotón y a todas las demás partículas neutras un valor de cero. La ley de conservación que se aplica a la carga eléctrica establece que el número cuántico total de carga eléctrica no puede cambiar en una interacción; la suma de todos los números cuánticos de carga después de un suceso debe ser igual a la suma antes del acontecimiento en cuestión.

Importa resaltar que la conservación de la carga eléctrica no prohíbe ningún cambio en el número de partículas cargadas. Un electrón y un positrón pueden aniquilarse entre sí, disminuyendo el número de partículas en dos: sin embargo, la carga total es cero tanto antes de la aniquilación como después. El proceso opuesto, en el que un electrón y un positrón se crean a partir de energía pura, obedece por la misma razón la ley de conservación. Cualquier partícula puede crearse o aniquilarse simultáneamente con su antipartícula y los números cuánticos serán conservados automáticamente.

El número cuántico llamado número bariónico es de notable interés para localizar la fuente de la asimetría cósmica entre materia y antimateria. Los bariones son una gran familia de partículas cuyos miembros más familiares son el protón y el neutrón; en su calidad de constituyentes básicos de los núcleos atómicos, los bariones tienen un papel claro e importante en la estructura de la materia ordinaria. El protón, el neutrón y todos los muchos bariones relacionados con ellos tienen asignado un número bariónico de + 1. Para el antiprotón, el antineutrón y los otros antibariones el número bariónico es − 1. Todas las demás partículas, incluyendo los piones, los muones, los neutrinos, el electrón, el fotón y sus antipartículas poseen un número bariónico igual a cero.

La conservación del número bariónico significa que, en cualquier reacción, el número bariónico de todas las partículas en el estado inicial es igual al número bariónico de todas las partículas en el estado final. De nuevo, el número de partículas puede cambiar; verbigracia: cuando un protón y un antiprotón se crean o aniquilan a pares; pero el número bariónico resultante permanece inalterado. Supongamos, por ejemplo, dos protones (con un número bariónico total de + 2) que chocan a alta energía. Los productos finales pueden incluir cuatro protones, un neutrón, tres antiprotones y un cierto número de piones; la suma de los números bariónicos muestra que el total sigue siendo + 2.

Se cree que la carga eléctrica es una cantidad que se conserva en todas las circunstancias. Menos cierta es la conservación absoluta del número bariónico. Existe de hecho una fuerte sospecha de que dicha ley se viola en algunas ocasiones.

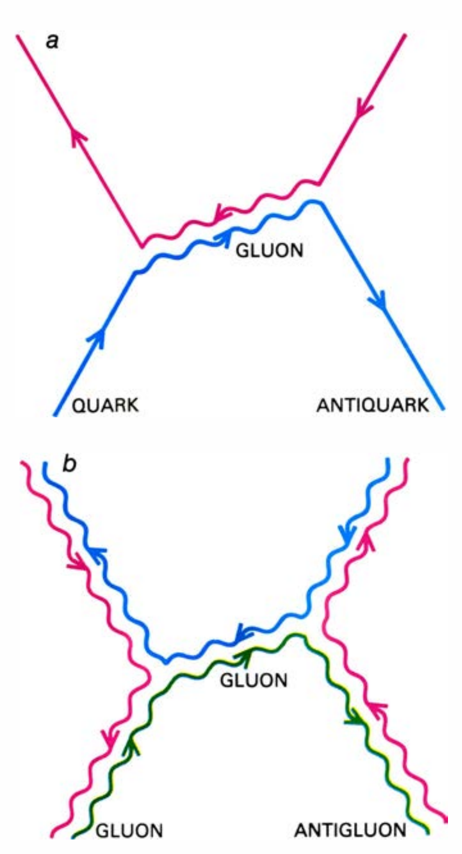
La prueba más concluyente de la conservación del número bariónico es la estabilidad del protón. Por tratarse de la partícula más ligera del grupo de las pesadas, cuyo número bariónico es + 1, el protón no puede desintegrarse en ningún conjunto de partículas más ligeras sin violar la ley de conservación. La detección de la desintegración de un protón constituiría, por tanto, una prueba directa de que dicha ley no se cumple siempre.

Nadie ha visto todavía desintegrarse un protón, y los mismos cálculos estimativos sugieren que posee una vida media larga. Por ejemplo, si los protones se desintegrasen en el esqueleto humano, la energía liberada aumentaría la incidencia de cáncer. Basándose en esto, la vida media del protón debería superar los 10^{16} años. Si los protones se desintegrasen en Júpiter, la energía contribuiría a la luminosidad del planeta. Sobre esta base, la vida media supera los 10^{18} años. Experimentos sistemáticos sugieren que la vida media rebasa, de hecho, los 10^{29} años. En contraste con ello, la edad del universo es sólo de 10^{10} años. Evidentemente, si el protón se desintegra, se trata de un suceso extremadamente raro. Si la vida media real oscila en torno a los 10^{30} años, en 100 toneladas de materia (un conjunto de 10^{31} protones) una media de 10 se desintegrarían en un año. Este bajo ritmo indica, a la vez, la severidad de la ley de conservación del número bariónico y la dificultad de realizar experimentos para buscar violaciones. No obstante, hay ya en marcha varios de tales experimentos.

Decir que el universo tiene un exceso de materia sobre antimateria equivale a decir que tiene un número bariónico positivo. Si la ley de conservación del número bariónico fuese absoluta, el número se hubiera mantenido constante a través de los tiempos. Cabría que en algún momento hubiera habido un número de bariones y antibariones mayor, pero el montante de bariones menos el de antibariones habría sido siempre el mismo.

Consideremos el estado del universo cuando tenía una centésima de segundo de edad y una temperatura de 10^{14} grados K. Para una temperatura dada existe una mezcla equilibrada de diferentes clases de partículas tal que, para cada clase, el número de partículas que se crean por colisiones o desintegraciones compensa el número que se destruyen. En el universo primitivo, a 10^{14} grados, la mezcla equilibrada comprendía unos 1000 millones de protones y unos 1000 millones de antiprotones por cada protón que hay en nuestro universo. Si el número bariónico del universo era el mismo que el de ahora, la razón de protones a antiprotones debió ser aproximadamente 1.000.000.001 a 1.000.000.000, de manera que la asimetría difícilmente sería observable.

Más tarde, casi todos los protones se aniquilaron en choques con antiprotones. Sólo la conservación del número bariónico evitó una aniquilación absoluta de todos los bariones y antibariones. En esta perspectiva, todos los protones presentes, y por tanto todas las galaxias, estrellas, planetas y seres vivos contemporáneos son los residuos de un desequilibrio de una parte en mil millones. Es este pequeño desequilibrio, la primitiva manifestación de la asimetría cósmica entre materia y antimateria, lo que precisa una explicación. Una vez que se hubo establecido el exceso de materia, la evolución subsiguiente del universo avanzó por una vía bastante directa; la fuente de la asimetría original constituye, empero, un misterio más profundo. En particular, si el universo evolucionó desde un estado inicial de completa simetría entre materia y antimateria (un estado que tenía número bariónico cero) hacia un estado asimétrico en el que el número bariónico era mayor que cero y el número de protones superaba al de antiprotones, entonces la conservación del número bariónico debió haberse violado en alguna etapa.



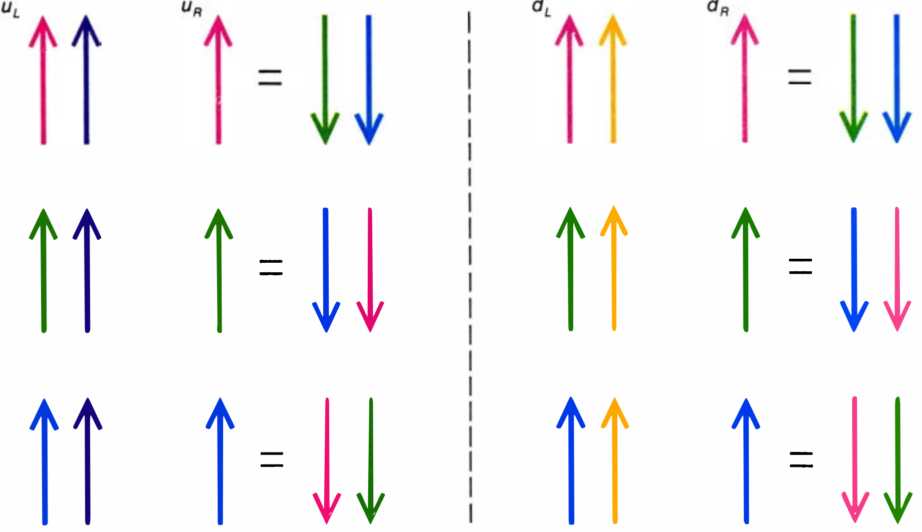
FUERZA FUERTE. Puede representarse como el intercambio de un gluon entre dos partículas que tienen la propiedad llamada color. En *a*, las partículas son quarks; el de la izquierda es azul, el de la derecha antiazul (nótese la dirección de la flecha). La interacción fuerte cambia la trayectoria y también el color de cada quark. Para conservar el color durante la interacción, el gluon debe tener a la vez un color y un anticolor; en virtud de lo cual los propios gluones están sujetos a la interacción fuerte. La colisión de un gluon y un antigluon se muestra en *b*. La interacción fuerte liga los protones y los neutrones en un núcleo atómico. La fuerza fuerte liga también los quarks que se cree que forman cada protón y neutrón.

La primera indicación de que la conservación del número bariónico no podía ser exacta vino de un campo de investigación muy alejado: la teoría del agujero negro. Un análisis desde el punto de vista matemático demostró que las únicas propiedades

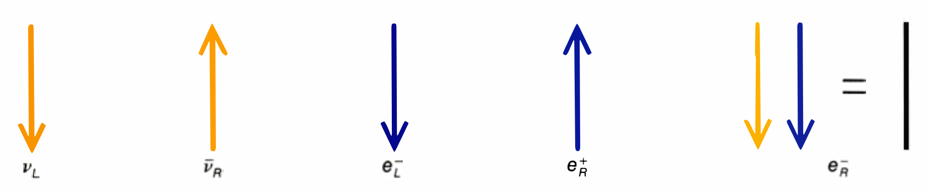
de un agujero negro medibles por un observador externo eran su masa, su momento angular y su carga eléctrica. Destaca la ausencia en esta lista del número bariónico. Por tanto, un agujero negro creado por el colapso de una estrella sería indistinguible de uno creado por el colapso de una antiestrella de la misma masa, momento angular y carga. Sin embargo, el número bariónico es positivo para la estrella mientras que es negativo para la antiestrella. Así pues, no hay manera de asignar un número bariónico a un agujero negro, y, por tanto, de estar seguro de que se conserve el número bariónico del universo.

La comúnmente aceptada violación de la ley de conservación por los agujeros negros sugiere que un mecanismo similar, en la escala microscópica, puede conducir a la desintegración del protón. En este hipotético proceso, un protón es absorbido por un agujero negro virtual: una pequeña fluctuación de corta vida en la geometría del espacio-tiempo, que, en principio, podría surgir en cualquier lugar y en cualquier instante. El agujero negro virtual rápidamente se desintegra en un positrón y un rayo gamma. En estas partículas reaparece la masa o energía del protón y también lo hace su carga eléctrica positiva; sin embargo, su número bariónico se pierde irremediablemente. Aunque los detalles del hipotético proceso son inciertos, las estimaciones sugieren que implica una vida media del protón del orden de 10^{40} años. Si la conservación del número bariónico se viola de esta manera, la violación es realmente tenue.

Una segunda indicación de que la conservación del número bariónico es tan sólo aproximada se nos muestra un tanto menos exótica y más potente también en su efecto sobre la vida media del protón. Este segundo mecanismo posible síguese de los procesos revolucionarios registrados en las teorías que describen las interacciones entre partículas elementales. En concreto, se trata de una consecuencia de la comprensión, lograda en la última década, de que la fuerza "fuerte", responsable de la estabilidad de los núcleos atómicos, y la fuerza "débil", responsable de la mayoría de desintegraciones radiactivas, son bastante similares a la fuerza electromagnética.



REPRESENTACION DE LA FAMILIA DE 12 QUARKS tal como aparecerían en diagramas de Feynman. Cada quark tiene cuatro atributos característicos. Primero, cada quark se llama arriba –up– (*u*) o abajo –down– (*d*). Los quarks *u* tienen una carga eléctrica de + 2/3, los quarks *d* tiene una carga eléctrica de − 1/3. (La carga de un electrón es − 1, la de un positrón o de un protón + 1.) Segundo, cada quark tiene un color con respecto a la interacción fuerte: rojo, verde o azul. Tercero, cada quark tiene un spin cuyo eje está alineado con la dirección de movimiento de la partícula (subíndice *R*, para partículas dextrógiras ["right-handed"]), o es opuesto a dicha dirección (subíndice *L*, para partículas levógiras ["left-handed"]). Finalmente, los quarks levógiros poseen un color con respecto a la fuerza de la naturaleza llamada interacción débil, que causa la mayoría de las desintegraciones radiactivas. El color es morado para los quarks *u* y naranja para los quarks *d*. Curiosamente, la fuerza débil no afecta a las partículas dextrógiras ni a las antipartículas levógiras. En el formalismo matemático de la interacción fuerte, un quark con un color dado es equivalente a un quark sin dicho color, pero con los otros dos anticolores. Se indica una equivalencia para cada quark dextrógiro. Una inversión de las flechas (que no se recoge en la representación esquemática superior) transforma un quark en el correspondiente antiquark.

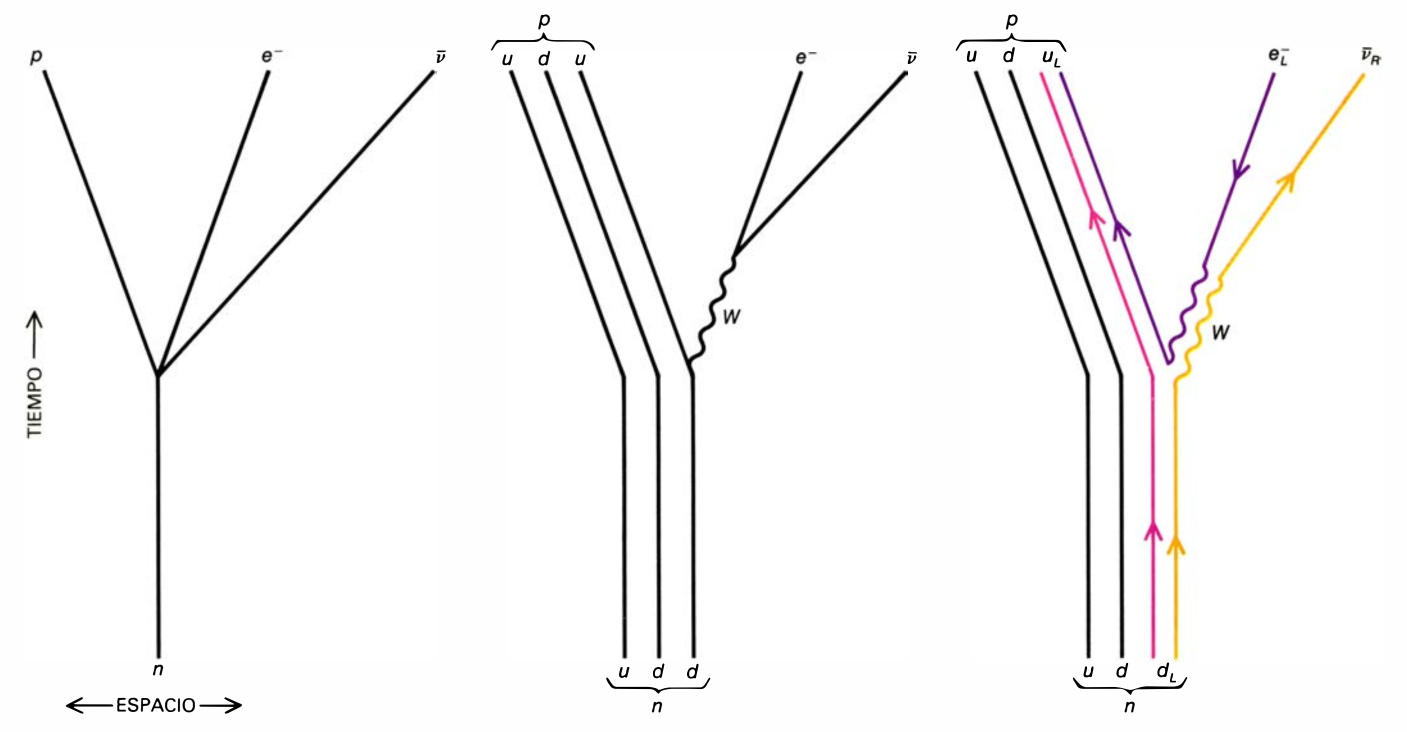


SON PARTICULAS DE INTERACCION DEBIL aquellas que incluyen el neutrino levógiro (subíndice *L*) y su antipartícula, el antineutrino dextrógiro (subíndice *R*), el electrón levógiro y su antipartícula, el positrón dextrógiro. Las partículas no interaccionan fuertemente, de manera que se muestran sin cargas de color fuerte. La quinta partícula es el electrón dextrógiro. Se da con dos cargas de color débil, una configuración que equivale a un estado privado de carga de color (*línea negra sin flecha de sentido*).

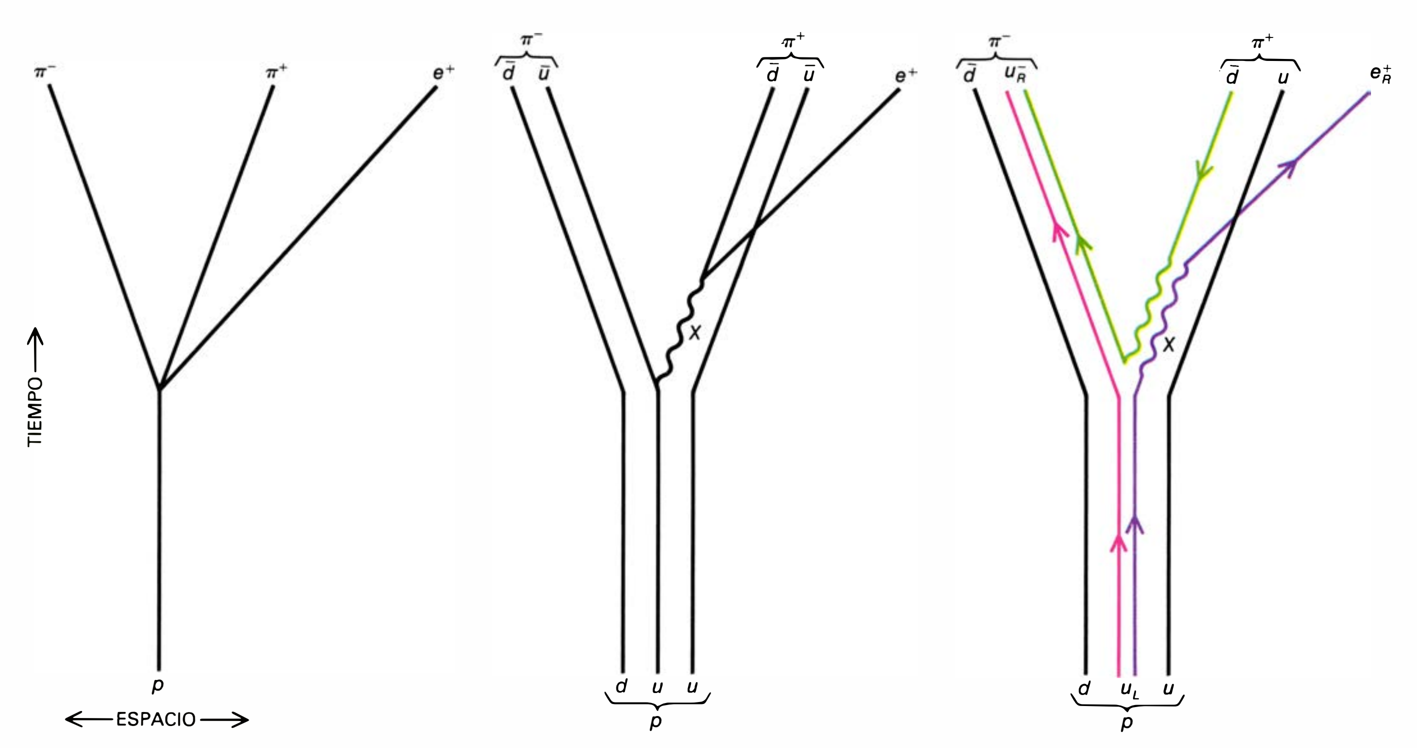
¿Cómo puede un avance en la comprensión de estas fuerzas, que no violan la conservación del número bariónico, conducir a teorías que predigan una tal violación? Un examen más pormenorizado de las distintas fuerzas debe preceder a la explicación.

De las tres fuerzas, sólo el electromagnetismo se pone en evidencia corrientemente en el mundo macroscópico que la gente percibe a su alrededor. La fuerza electromagnética actúa sólo entre partículas que tienen carga eléctrica; la interacción puede describirse como el intercambio de una tercera partícula, un fotón. Se dice que el fotón es una partícula vectorial, denominación que reciben las partículas cuyo momento angular de spin, cuando se mide en unidades fundamentales, es igual a 1. Quizá la característica más fundamental del electromagnetismo sea que puede describirse por una teoría invariante bajo transformaciones de "gauge" o de contraste. En una teoría de esta clase, el origen de la fuerza tiene que ver con una ley de conservación, en este caso la conservación de la carga eléctrica. El acoplamiento de las partículas vectoriales a una carga conservada es característico de las teorías invariantes de contraste.

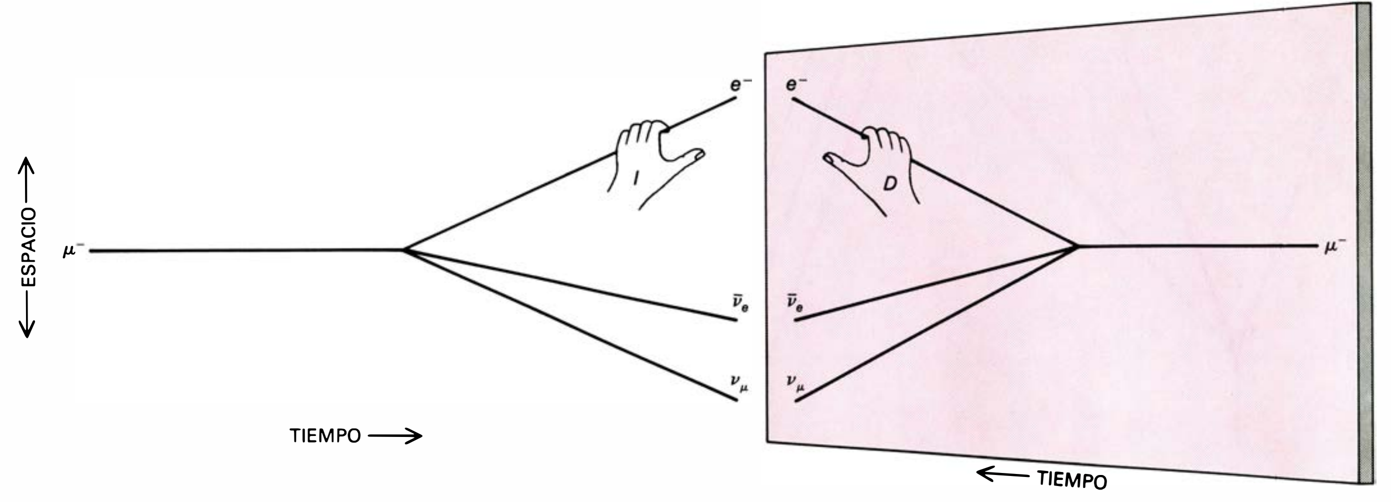
En todos estos aspectos, la interacción fuerte se comporta de un modo parecido. La fuerza surge de una teoría de contraste y la interacción fuerte puede describirse como el intercambio de una partícula vectorial entre dos otras partículas que tienen una cierta clase de carga. Sin embargo, la partícula vectorial no es aquí el fotón, sino un ente hipotético llamado gluon, y la carga no es la carga eléctrica, sino una propiedad llamada color. La carga de color no tiene nada que ver, desde luego, con el color en el sentido ordinario. En este contexto, la palabra carga es menos fantasiosa. La palabra es adecuada porque la carga de color desempeña el mismo papel en la interacción fuerte que el que representa la



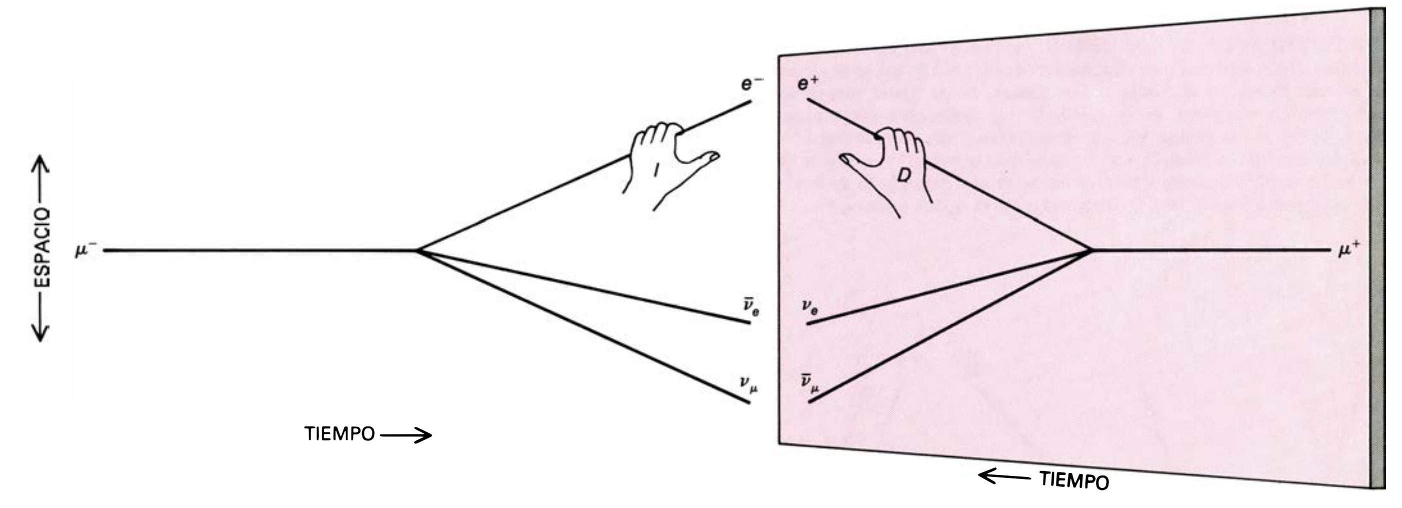
DESINTEGRACION DEL NEUTRON. Se trata de una interacción débil que permite a las partículas una vida media de sólo unos 15 minutos en promedio si no está ligada en el núcleo de un átomo. En la figura menos detallada (*diagrama de Feynman de la izquierda*), la desintegración transforma el neutrón (*n*) en un protón (*p*), un electrón (*e*) y un antineutrino ($\bar{\nu}$). En un análisis más fino (*centro*), el neutrón consta de tres quarks, y sólo un quark *d*, colocado arbitrariamente a la derecha, se ve afectado por la desintegración. Un quark *d* levógiro (d_L) se desintegra en el quark *u* levógiro (u_L), más ligero, el electrón y el antineutrino. El quark d_L era rojo respecto a la interacción fuerte y naranja con respecto a la débil. La carga naranja se transforma por una partícula (*W*) de vida corta (o virtual), que es naranja y antimorada. Sólo los colores débiles se cambian. (Las flechas indican la dirección de flujo de las cargas de color.) Una interacción débil de la clase aquí representada no sólo es típica de la desintegración del neutrón, sino también de la mayoría de las otras clases de desintegración radiactiva. Los colores de los quarks que no cambian no se representan en el diagrama de la derecha.



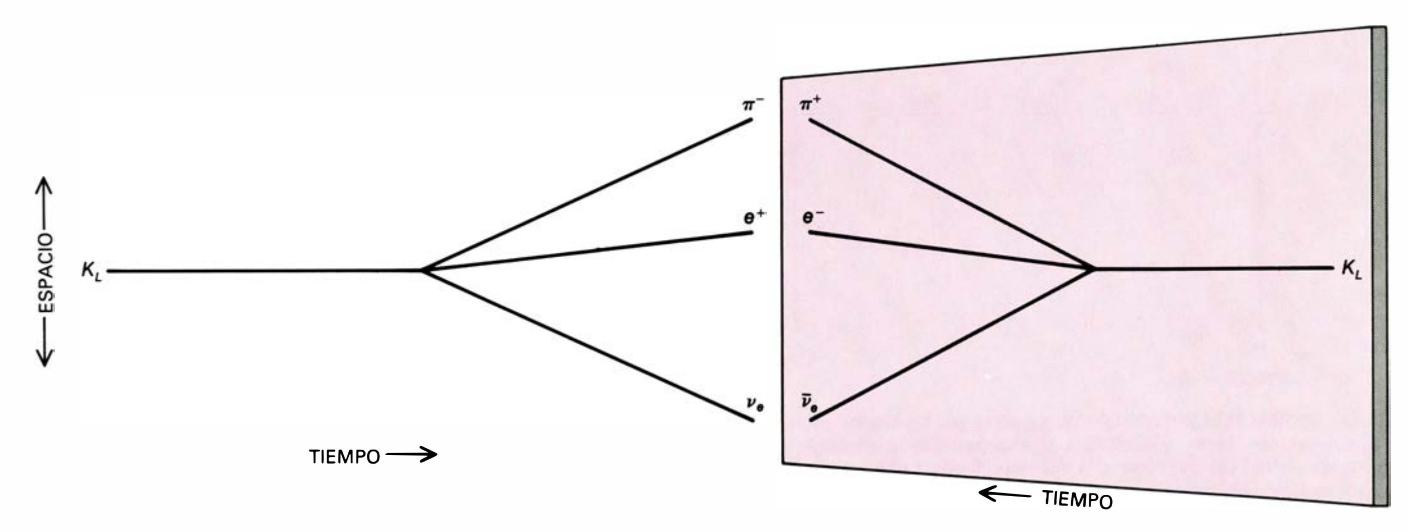
DESINTEGRACION DEL PROTON, predicha por las teorías que unifican las interacciones fuerte y débil. Las teorías postulan la existencia de una partícula virtual (*X*) que tiene a la vez color fuerte y débil y que causa la desintegración. Su creación espontánea es rara, por lo que se cree que el protón sobrevive, en promedio, unos 10^{31} años. En el modo de desintegración mostrado, el protón (*p*) se transforma en dos piones cargados (π^+ y π^-) y un positrón (e^+). El suceso microscópico responsable de la desintegración es la transformación de un único quark en dos antiquarks y el positrón. De nuevo, el diagrama de la izquierda muestra el esquema de la desintegración, el diagrama del centro descompone las partículas en quarks y el de la derecha señala el flujo de cargas de color para el quark que se transforma. El protón es el miembro más ligero de la clase de partículas llamadas bariones. Si el protón puede desintegrarse, no será ley de la naturaleza la constancia del número de bariones del universo. Esta ley de la constancia debió infringirse en algún momento de la historia del universo, caso de que se desarrollara una asimetría entre materia y antimateria en el universo primitivo.



LA VIOLACION de la conservación de la paridad fue el primero de una serie de descubrimientos que mostraron que las leyes de la física no eran indiferentes a la distinción entre materia y antimateria. La conservación de la paridad, P, mantiene que cada proceso físico permanecería invariante si pudiera transformarse en su imagen especular. El proceso representado es la desintegración de un muon (μ) en un electrón (e^-), un antineutrino electrónico ($\bar{\nu}_e$) y un neutrino muónico (ν_μ). El electrón es levógiro (*letra L*). En la reflexión especular de la desintegración, el electrón es dextrógiro (*letra R*). De hecho, la paridad no se conserva en este proceso: los electrones levógiros aparecen más de 1000 veces más a menudo que los dextrógiros.



LA CONSERVACION DE CP proponía una simetría que podía observarse aun cuando se violase la conservación de la paridad. La conservación de CP establece que la simetría rota por una reflexión especular puede restablecerse sustituyendo todas las partículas por sus antipartículas. En la desintegración del muon, la conservación de CP se mantiene: la desintegración de la izquierda y la desintegración de la derecha resultan ser igualmente comunes.



VIOLACION DE CP: se demostró en la desintegración del mesón *K* neutro de vida larga (K_L). La desintegración de esta partícula en π^-, e^+ y ν_e es más corriente que su propia desintegración en las antipartículas π^+, e^- y $\bar{\nu}_e$. (El mesón *K* neutro de vida larga coincide con su antipartícula.) Si la simetría CP no se rompiera en ningún caso, la razón de bariones a antibariones sería fija y no se podría desarrollar ninguna asimetría entre materia y antimateria.

carga eléctrica en la interacción electromagnética.

Una diferencia entre el electromagnetismo y la interacción fuerte consiste en que el electromagnetismo tiene sólo una clase de carga, mientras que en la interacción fuerte hay tres, llamadas *R, G* y *B* por rojo ("red"), verde ("green") y azul ("blue"). Los constituyentes fundamentales de todas las partículas que interaccionan fuertemente, los quarks, tienen color. Cada quark posee un color único, designado por un valor de los tres números cuánticos de color. Para los quarks rojos, *R* es igual a + 1 mientras que *G* y *B* son iguales a 0. Similarmente para los quarks verdes, *G* vale + 1 y, para los quarks azules, *B* equivale a + 1 y los otros números cuánticos de color son cero. La teoría exige ocho clases de gluones. Seis clases cambian un quark de un color en un quark de un color distinto de todas las maneras posibles, a saber: rojo en verde, rojo en azul, verde en rojo, verde en azul, azul en rojo y azul en verde. Los otros dos gluones se parecen al fotón en que llevan una fuerza entre partículas "cargadas", pero no alteran la carga.

Propiedad de las cargas de color es que pueden cancelarse entre sí. Así, la combinación de un quark rojo, otro verde y un tercero azul es una partícula compuesta sin color a la que los gluones no se acoplan. (De manera similar, partículas con cargas eléctricas opuestas pueden combinarse para formar un compuesto neutro.) Solamente estas combinaciones incoloras de quarks parecen existir en la naturaleza. Todos los bariones están formados por tres quarks, uno de cada color. Los mesones, otra categoría de partículas que interaccionan fuertemente, constan cada uno de un quark y un antiquark.

Una segunda diferencia entre la interacción fuerte y el electromagnetismo es que los propios gluones están cargados, en tanto que el fotón no lo está. Por ejemplo, el gluon que es absorbido por un quark rojo y lo transforma en un quark verde tiene *R* igual a −1, *G* igual a +1 y *B* igual a 0; con esta combinación de colores y anticolores la carga de color se conserva durante la interacción. Puesto que los gluones se acoplan a las partículas coloreadas, y siendo coloreados los propios gluones, éstos se acoplan entre sí. Por el contrario, el fotón es eléctricamente neutro y no se acopla a otros fotones. Esa diferencia comporta una profunda consecuencia dinámica: a cortas distancias, la interacción fuerte pierde intensidad. Los quarks se ligan sólo débilmente si se hallan muy juntos, pero su enlace se hace intenso cuando media cierta distancia. (En este contexto, una gran distancia son 10^{-13} centímetros.)

Esta paradójica ley de fuerzas explica mucho. Desde mediada la década de 1960 se sabe que las propiedades de las partículas que interaccionan fuertemente podían explicarse a través del modelo de los quarks, pero nadie ha observado nunca un quark aislado. Además, la utilidad de tratar una partícula que interacciona fuertemente como un compuesto de quarks se apoya en una aproximación según la cual los quarks son esencialmente partículas que no interaccionan dentro de una "bolsa" ("bag") común. No dejaba de sorprender que partículas que interaccionan fuertemente, los quarks por ejemplo, pudieran describirse con éxito cual si se tratara de no interaccionantes. La idea de que la intensidad de la interacción fuerte entre los quarks decrece cuando los quarks están juntos explica de manera clara por qué los quarks confinados dentro de una "bolsa" interaccionan sólo débilmente entre sí, por más que no puedan ser separados. Quizá resulte imposible aislar un quark. La teoría de contraste de la interacción fuerte que sustenta el modelo de quark conduce a muchas predicciones experimentales, hasta la fecha muy acertadas. La teoría está ganando aceptación casi universal.

Podemos describir la interacción débil de la misma manera que las interacciones electromagnéticas y fuertes, con la salvedad de algunos rasgos propios. En primer lugar, hay dos clases de carga análogas a las tres cargas de color de la interacción fuerte. Les llamaré *P* y *O*, de los colores morado ("purple") y naranja ("orange"). Tres partículas vectoriales, llamadas W^+, W^- y *Z*, transportan la interacción. Estas partículas tienen masas grandes, al contrario del fotón y los gluones, que carecen de masa. Una partícula de gran masa puede surgir espontáneamente sólo como una fluctuación de vida corta; si tiene vida corta, no puede ir lejos y, en consecuencia, la interacción débil goza de un alcance muy limitado. Una característica más sorprendente de la fuerza débil es que sólo actúa sobre partículas dotadas de ciertas propiedades geométricas. Los quarks, electrones, neutrinos y unas partículas más, no muchas, pueden clasificarse como dextrógiras o levógiras según sea la orientación relativa de su momento angular de spin y su momento lineal. La partícula dextrógira tiene su eje de spin paralelo a su dirección de movimiento, y, antiparalelo, una partícula levógira. La interacción débil afecta sólo a las partículas levógiras y a las antipartículas dextrógiras. En resumen, las interacciones fuertes y débiles requieren cinco clases de carga de color (rojo, verde y azul para la fuerte y morado y naranja para la fuerza débil) junto con partículas vectoriales que transforman algunos de estos colores.

En las teorías que he indicado aquí, la fuerza fuerte constituye un mecanismo para cambiar los colores rojo, verde y azul de los quarks. La fuerza débil efectúa cambios similares en los números cuánticos de color morado y naranja de las partículas. Si estas teorías han de engarzarse de una manera verdaderamente unificada, cabría esperar alguna fuerza adicional que transformara los colores fuertes en los colores débiles, y viceversa. Además de su elegancia conceptual, un esquema que incorpore esa nueva fuerza encasillará de un modo coherente todas las partículas conocidas. Por si ello fuera poco, hace predicciones definidas. Así, pronostica la masa de la *W*, una partícula todavía por detectar.

Postulando una nueva fuerza es como las teorías unificadas ponen en entredicho la conservación del número bariónico y permiten que el protón se desintegre. Se introducen nuevas partículas vectoriales que cambian el color a modo de puentes entre las partículas con color fuerte, tales como los quarks que componen el protón, y partículas con sólo color débil, cuyo número bariónico es cero. Yo designaré por *X* tales partículas vectoriales. La teoría unificada predice que la *X* tiene una masa que es 10^{15} veces mayor que la masa del protón (lo que viene a significar la masa de una pulga), comprimida en un volumen de unos 10^{-27} centímetros de diámetro. Por ser tan pesada la partícula *X*, su creación espontánea resulta extremadamente rara. En consecuencia, se estima que la vida media del protón, aunque larga, no es infinita; la vida media debe ser del orden de 10^{31} años.

Una vida media de 10^{31} años implica que, en el universo actual, la violación de la conservación del número bariónico es pequeña. Sin embargo, como he indicado, la asimetría materia-antimateria observada hoy día corresponde simplemente a una asimetría de una parte en mil millones en el universo primitivo. Además, un modo de desintegración que exige la creación de una

partícula pesada inestable quizá fuera más corriente en los primeros momentos del universo, cuando las partículas pesadas podían crearse libremente en choques a energías ultraelevadas.

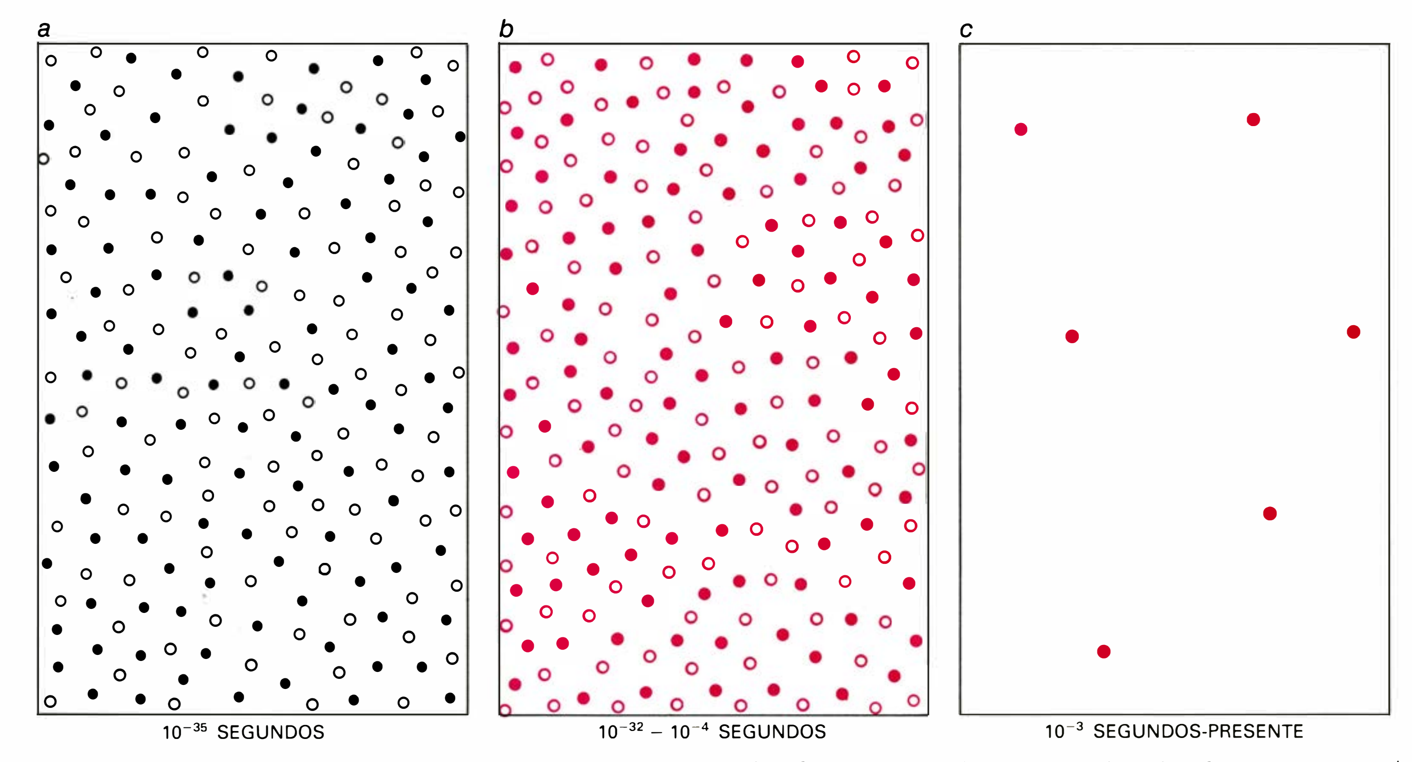
Volvamos a la idea de que las leyes físicas son indiferentes a la distinción entre materia y antimateria. La historia de la idea está jalonada de sorpresas. Hasta la mitad de la década de 1950 se aceptaba que las leyes físicas permanecerían invariantes si los experimentos se repitiesen en un mundo reflejado en un espejo. En otras palabras, se creía que no había lugar a una distinción absoluta entre izquierda y derecha. Sin embargo, varios experimentos mostraron entonces que la simetría de reflexión en un espejo se rompía completamente por las interacciones débiles. Un ejemplo lo proporcionó la desintegración del muon en un electrón, un neutrino y un antineutrino. De 1000 desintegraciones, en más de 999 se observó que el electrón era levógiro: su eje de spin apuntaba en una dirección opuesta a su dirección de movimiento. De este modo, la desintegración del muon proporcionaba un patrón absoluto de la izquierda frente a la derecha.

Los teóricos propusieron, a renglón seguido, una simetría más generalizada que parecía iban a respetar todas las interacciones. Esta segunda hipótesis postulaba que las leyes de la física serían invariantes por la reflexión en un espejo de un experimento si, al mismo tiempo, se sustituían todas las partículas del experimento por sus antipartículas. La simetría se llama CP, por conjugación de carga y paridad, o reflexión especular. La simetría CP predice que, en la desintegración del antimuón, emergerá su positrón en vez de un electrón, y que el positrón debe ser casi siempre dextrógiro. En el caso de la desintegración del muon se observa una simetría CP exacta.

Si la simetría CP fuese absoluta, no podría generarse una preponderancia de materia o de antimateria a partir de una igualdad primordial entre materia y antimateria. Para cada proceso reflejado e igualmente probable crearía la antipartícula.

La idea de absoluta simetría CP se mantuvo durante unos siete años. Pero se observó que el mesón K neutro de vida larga, que coincide con su propia antipartícula, se desintegra más a menudo en un pion negativo, un positrón y un neutrino, que en un pion positivo, un electrón y un antineutrino. Si CP fuera una simetría absoluta, los dos modos de desintegración deberían ser igualmente probables. No se ha encontrado ninguna otra violación de la simetría CP distinta de la desintegración del mesón K, pero tales violaciones pudieran tener un papel más importante en la naturaleza a energías ultraaltas.

De cuanto se ha expuesto se deduce que la permanencia de ciertas partículas, tal como se formaliza en la ley de conservación del número bariónico, y la indiferencia de las leyes físicas para distinguir entre materia y antimateria, tal como se formaliza en el principio de la simetría CP, no son exactas, sino sólo aproximadas. Es cierto que los principios se cumplen con bastante exactitud en la actualidad, pero cabe que no hubiera sido así en los albores del universo. De hecho, dada una pequeña violación de estos principios, uno puede construir una cadena concreta de sucesos que arrancan de un universo con un estado inicial simétrico entre materia y antimateria y llegan a un universo donde rige una preponderancia de materia sobre antimateria.



EVOLUCION DE LA ASIMETRIA COSMICA entre materia y antimateria. Se representa de acuerdo con las predicciones de las teorías que unifican las interacciones fuerte y débil. El panel *a* simboliza el universo 10^{-35} segundos después de la gran explosión. El panel muestra iguales cantidades de partículas *X* (*puntos negros*) y de sus antipartículas, $\overline{X}$ (*círculos negros vacíos*). En el recién nacido universo, tales partículas se producían copiosamente en colisiones de energía ultraelevada. El panel *b* muestra el universo desde los 10^{-34} segundos hasta los 10^{-4} segundos. Las *X* y las $\overline{X}$ se han desintegrado en formas que no siempre conservan el número bariónico y la simetría CP. El resultado es un ligero desequilibrio que prima a los protones (*puntos coloreados*) sobre los antiprotones (*círculos coloreados vacíos*). El panel muestra seis protones más que antiprotones; el desequilibrio era, de hecho, mucho menor, concretamente de una parte de mil millones. Aunque el protón es menos pesado que la *X* en un factor de 10^{15}, el protón se representa mayor porque, de acuerdo con la teoría de la mecánica cuántica, la masa es inversamente proporcional a la incertidumbre en la posición. El panel *c* indica el universo desde 10^{-3} segundos de vida hasta ahora. Cada encuentro de un protón y un antiprotón ha causado la aniquilación de ambas partículas y sólo sobreviven los protones de más. Contribuyen a la preponderancia manifiesta de materia sobre antimateria observada en el universo actual.

La cadena de razonamientos se inicia con la observación de que la temperatura del universo ha estado bajando constantemente desde la gran explosión. Cuanto más alta era la temperatura, tanto mayor era la velocidad media y la energía de las partículas que formaban el universo y, por ende, mayor resultaba la energía disponible en una colisión para la creación de otras partículas. A una temperatura superior a los 10^{28} grados K, la energía típica de una partícula era comparable a la energía de la masa en reposo de una partícula X. Hasta unos 10^{-35} segundos después de la gran explosión, el universo tuvo una tal temperatura; puede postularse, pues, que existía una gran densidad de partículas X.

A medida que el universo se expandía y se enfriaba, la probabilidad de crear una partícula X bajó rápidamente; mientras tanto, las partículas existentes se desintegraron en seguida. Supongamos que las desintegraciones no conserven el número bariónico. Una partícula X puede desintegrarse entonces en uno de entre varios estados finales con distinto número bariónico total. La media puede ser, digamos, de $+ 2/3$. Si el universo tenía iguales cantidades de materia y antimateria antes de cumplir los 10^{-35} segundos de edad, incluiría un número igual de X y de $\overline{X}$, donde la $\overline{X}$ es la antipartícula de la X. Puede parecer, por tanto, que cada modo de desintegración de una X estaría compensado por la desintegración de una $\overline{X}$, que daría partículas con un número bariónico medio de $-2/3$. En tal caso, el número bariónico total del universo permanecería siempre nulo. De hecho, como la simetría CP puede no haberse cumplido exactamente en la desintegración de la X y la $\overline{X}$, no cabe concluir que las dos series de desintegraciones siempre den conjuntos de partículas simétricamente opuestos. La $\overline{X}$ puede generar partículas cuyo número bariónico medio no fuera $-2/3$, sino, por ejemplo, $-1/3$.

De esta manera, un universo que tuviera el mismo número de partículas X y $\overline{X}$ podría haberse convertido en un universo con un número bariónico positivo y una correspondiente preponderancia de materia. Podría haber sido, por ejemplo, un universo con un desequilibrio de una parte en mil millones a favor de la materia. Transcurridos unos 10^{-35} segundos, la temperatura y la energía típica por partícula en todo el universo habría caído debajo del umbral para la creación de una X y una $\overline{X}$. Los procesos que violan el número bariónico habrían pasado a ser insignificantes y la preponderancia de materia sobre antimateria habría quedado estable. El universo tendría muchos más bariones y antibariones de los que tiene ahora, pero la mayoría de ellos se aniquilarían entre sí, dejando el residuo de materia que se observa hoy en día.

Varios puntos de este razonamiento argumentativo son de orden altamente especulativo; y quizás la explicación de la asimetría cósmica entre materia y antimateria dé la impresión de pertenecer a un terreno más mítico que científico. Hasta cierto punto, esto resulta inevitable, ya que las condiciones extremas del universo primitivo no pueden reproducirse en un laboratorio. Lo que distingue la especulación científica del mito es su coherencia lógica y la comprobación de algunos, al menos, de sus elementos por una prueba experimental. He descrito cómo la lógica interna de la física de partículas conduce a las teorías unificadas en las que el número bariónico no se conserva y he indicado que futuros desarrollos tanto en astronomía de neutrinos como en la búsqueda de la desintegración del protón someterán las teorías a prueba. Si estos difíciles experimentos dan resultados concordes con las expectativas teóricas, avanzaremos mucho en la comprensión científica de una simetría misteriosa. Ya ahora, cálculos realizados de acuerdo con las teorías unificadas sugieren que la densidad media de materia en nuestro universo es coherente con el transcurso primordial de los sucesos que las teorías unificadas implican. Debido a las incertidumbres que rodean a los mecanismos de violación de CP, resulta difícil realizar los cálculos con precisión, pero el cuadro cualitativo que se desprende parece convincente.

Queda otro problema. He descrito cómo pudo haber empezado el universo en un estado de asimetría entre materia y antimateria y haberse desarrollado luego de manera asimétrica. ¿Por qué esa simetría en un principio?

A un cierto nivel, esta pregunta puede contestarse estadísticamente. Aun cuando las interacciones que violan el número bariónico fueron frecuentes en el universo primitivo, la condición universal más probable que sería alcanzada en equilibrio antes de 10^{-35} segundos es aquella en que la cantidad de materia iguala la cantidad de antimateria. Las teorías unificadas imponen, por tanto, la simetría inicial automáticamente; no precisa ser postulada por separado. Después de 10^{-35} segundos, los ritmos de desintegración de las X y las $\overline{X}$ habrían sido lentos en comparación con los ritmos de expansión y enfriamiento del universo. Bajo tal condición, ya no se puede alcanzar el equilibrio.

Pero a un nivel más profundo, no acaba de convencerme dicha explicación. No logra justificar por qué el universo debió de haber empezado con un suceso explosivo. Tampoco razona por qué el universo es simétrico en varios otros aspectos: es eléctricamente neutro en promedio y parece no tener un momento angular global. Expondré una idea que tal vez ayude a entender esas cuestiones. No está en absoluto bien establecida, pero sugiere un programa de investigación. En realidad, fue la que me indujo a ahondar en mis trabajos sobre la asimetría materia-antimateria.

Las teorías modernas de las interacciones entre partículas elementales ponen de manifiesto que el universo puede existir en diferentes fases, en cierto modo análogas a las fases sólida y líquida del agua. En fases distintas, las propiedades de la materia difieren; por ejemplo, una cierta partícula puede no tener masa en una fase, pero poseerla en otra. Las leyes de la física son más simétricas en algunas fases que en otras, de la misma forma que el agua líquida es más simétrica que el hielo, donde la red cristalina distingue ciertas posiciones y direcciones del espacio.

En estas teorías, la fase más simétrica del universo resulta ser en general inestable. Uno puede especular que el universo empezó en el estado más simétrico posible y que en tal estado no existía materia; el universo era un vacío. Un segundo estado era posible y en él existía materia. El segundo estado, ligeramente menos simétrico, también poseía menos energía. Andando el tiempo, apareció una zona de la fase menos simétrica y creció rápidamente. La energía liberada por la transición tomó forma en la creación de partículas. Este suceso podría identificarse con la gran explosión. La neutralidad eléctrica del universo de partículas estaría así garantizada, porque el universo sin materia había sido eléctricamente neutro. La falta de rotación en el universo de materia podría entenderse como una de las condiciones más favorables para el cambio de fase y el subsiguiente crecimiento, con todo lo que el crecimiento implicaba, incluyendo la asimetría cósmica entre materia y antimateria. La respuesta a la vieja pregunta "¿Por qué hay algo en vez de nada?" sería: porque "nada" es inestable.

Ciencia y sociedad

El tercer carbono

El carbono genera más diversidad estructural en las moléculas que el resto de elementos químicos tomados en su conjunto. A pesar de ello, durante mucho tiempo se consideró que, cuando se combinaba consigo mismo, sólo producía dos tipos de cristal. En el diamante, que se forma bajo temperatura y presiones elevadas, cada átomo de carbono está unido mediante un enlace sencillo con otros cuatro átomos vecinos, formando una red tridimensional. Cada enlace supone compartir un par de electrones. En el grafito, producido a temperaturas y presiones bajas, los átomos de carbono se disponen en planos formados por anillos de seis carbonos, donde los enlaces sencillos alternan con los dobles. Cada doble enlace supone la compartición de dos pares de electrones. No se comparten electrones entre un plano y otro, lo que explica que la red cristalina del grafito sea débil, y el propio grafito muy blando.

En 1968 se descubrió, casi simultáneamente en los Estados Unidos y en la Unión Soviética, una tercera forma cristalina del carbono. Aunque no se ha logrado desentrañar los pormenores de su estructura cristalina, parece que se trata de un conjunto de formas donde se unen los átomos de carbono a través de una mezcla de enlaces sencillos y triples. Por cada triple enlace se comparten tres pares de electrones.

Las nuevas formas cristalinas han sido denominadas carbino, pues el sufijo *-ino* representa convencionalmente al átomo de carbono unido mediante triple enlace. La gama de estos cristales se extiende desde formas blandas hasta formas superduras, o casi tan duras como el diamante. Tienden a condensarse a partir de carbono gaseoso a altas temperaturas y bajas presiones. La presión puede, no obstante, acercarse a los 70.000 kilogramos por centímetro cuadrado, presión umbral de la formación de diamante. Se ha encontrado carbino mezclado con grafito en varios yacimientos naturales; también ha podido sintetizarse en el laboratorio. Se sabe que forma parte de los meteoritos y del polvo del espacio interestelar. Lejos de ser un compuesto raro, su abundancia pudiera equipararse, en el universo, con la del grafito.

La existencia de carbino en el espacio se aventuró hace dos años, cuando Adrian S. Webster, de la Universidad de Cambridge, interpretó que ciertas características de la radiación infraroja espacial se debían a la emisión de moléculas que contenían una cadena de nueve a 11 átomos de carbono unidas, alternativamente, por enlaces simples y triples. Cada molécula tenía sólo un átomo de hidrógeno. No resultaba fácil imaginar cómo una molécula tan compleja y pobre en hidrógeno, elemento predominante en el universo, podía haberse formado en el espacio. Webster sugirió que podía tratarse de un fragmento desprendido de un cristal de carbino.

La bibliografía científica cuenta con muchos trabajos que abordan la presencia de carbino en los meteoritos Allende y Murchison. Ambos son ricos en carbono, y se supone que los dos son antiguos: aparentemente no han sufrido alteraciones por fusión o colisión desde la formación del sistema solar. Se han utilizado tres técnicas distintas para tratar de revelar la presencia de carbino en los meteoritos. En una de ellas, aplicada por A. Greenville Whittaker y sus colegas, de la Aerospace Corporation, se dispersan los electrones de los cristales contenidos en muestras de meteoritos; los ángulos de dispersión indican las posiciones de los átomos en los cristales. Los resultados sugieren que algunas zonas de las muestras presentan la estructura propuesta para el carbino. También se ha empleado este método en la detección de pequeñas regiones de carbino contenidas en grafito terrestre.

Las otras dos técnicas detectan el carbino de forma más directa. Las está aplicando el grupo de Whittaker y Ryoichi Hayatsu, Robert G. Scott y Martin H. Studier, del Argonne National Laboratory, y Roy S. Lewis y Edward Anders, de la Universidad de Chicago. Se liberan fragmentos moleculares de los meteoritos por medio del calor o por bombardeo con un haz de partículas. Se descubrió, durante los primeros segundos, que en los fragmentos abundaban los compuestos con una masa múltiplo de la masa de un átomo de carbono. Ello implicaba que los fragmentos eran cadenas de átomos de carbono. Los múltiplos pares superaban en número a los impares, lo que se tomó como prueba de la alternancia de enlaces simples y triples. Los enlaces triples son más resistentes a la ruptura. Si el carbono estuviera en forma de grafito, abundarían los fragmentos seis veces más pesados que el átomo de carbono. Si se tratara de carbono orgánico, se habría detectado hidrógeno y los fragmentos tendrían pesos distintos de los que poseen los átomos de carbono aisladamente considerados.

Se ha descubierto recientemente que el carbino de los meteoritos Allende y Murchison ha absorbido trazas de gases inertes, de entre los cuales destaca especialmente el xenón. El proceso habitual de acumulación de gases es la rotura de núcleos pesados, bien por desintegración radiactiva, bien a consecuencia del impacto de un rayo cósmico de alta energía. Según la hipótesis de que la nube de gases cuya condensación formó el sistema solar era homogénea, las proporciones de los diversos isótopos de un gas inerte serían las mismas en cualquier muestra de materia del sistema solar. En caso contrario, una mezcla no usual de isótopos sugeriría dos alternativas: el material que contiene los isótopos no es representativo del conjunto del sistema solar, o puede que proceda de algún otro lugar.

En el carbino del meteorito de Allende se encontraron los nueve isótopos del xenón, en una concentración excepcionalmente elevada. El enriquecimiento en los isótopos más pesados supone que derivan de un elemento extremadamente pesado. Sin embargo, no hay ningún elemento en el meteorito de Allende que reúna estas características. Se conjetura que pudiera tratarse de un elemento superpesado, ausente hoy día del sistema solar.

El origen del xenón del carbino del meteorito de Murchison quizá sea aún más extravagante. Las concentraciones relativas de los isótopos del xenón sugieren una desintegración radiactiva de núcleos formados por una lenta captura de neutrones. La distribución de concentraciones corresponde en concreto a los modelos de captura lenta de neutrones de las estrellas gigantes rojas, que, habiendo agotado su hidrógeno, brillan en virtud de la fusión de helio. Al cambiar de combustible nuclear, aparece en estas estrellas una extensa envoltura gaseosa.

La historia del carbino del meteorito de Murchison se remontaría por tanto a la atmósfera de una estrella gigante roja, donde se habría condensado a partir de carbono gaseoso hace varios miles de millones de años. El carbino podría haber capturado trazas del material de la estrella. Esta materia se

habría sintetizado, en parte, por captura lenta de neutrones, y, en el transcurso del tiempo, la desintegración radiactiva de la materia habría dado lugar a los isótopos del xenón. El carbino habría ido a parar a la nube de gas que, por condensación, formó el sistema solar, y se incorporó a una matriz que fue a caer a la tierra.

Sanidad a paso lento

El rápido avance de los conocimientos biomédicos y de la tecnología asociada registrado en los últimos años no se ha visto acompañado de una mejoría de la salud mundial. Hay una brecha, quizá cada vez más abierta, entre las posibilidades médicas, de un lado, y la prevención y cura de las enfermedades, de otro. En los países desarrollados, igual que en los subdesarrollados, la pobreza sigue siendo la raíz de los problemas sanitarios más apremiantes. El enfoque tradicional de la mejora de la asistencia médica, que hace hincapié en el desarrollo de una tecnología dirigida a la curación de enfermedades agudas de pacientes selectos, no logra resolver los problemas sanitarios que atañen a las masas de los países pobres y está resultando demasiado caro para los ricos. Son estas algunas de las conclusiones que se recogen en el *Sexto Informe sobre la Situación Sanitaria Mundial,* elaborado por la Organización Mundial de la Salud, y que abarca el período que va de 1973 a 1977.

A quienes están al corriente del descenso de las enfermedades infecciosas en las naciones industrializadas y del papel preeminente que están tomando las enfermedades cardíacas y el cáncer, les ha cogido por sorpresa que "la mayoría de las muertes del mundo sigan debiéndose a la misma mezcla de enfermedades infecciosas, parasitarias y respiratorias, unidas a carencias nutritivas". Salvo el éxito habido en la erradicación de la viruela, el progreso en la reducción de las enfermedades que afligen con mayor gravedad a los países subdesarrollados, en los que habitan las tres cuartas partes de la población mundial, ha sido "escaso o nulo". La incidencia de la mayoría de estas enfermedades "podría reducirse espectacularmente a un costo relativamente bajo", de forma que "el sufrimiento y la muerte que generan han de considerarse innecesarios y evitables".

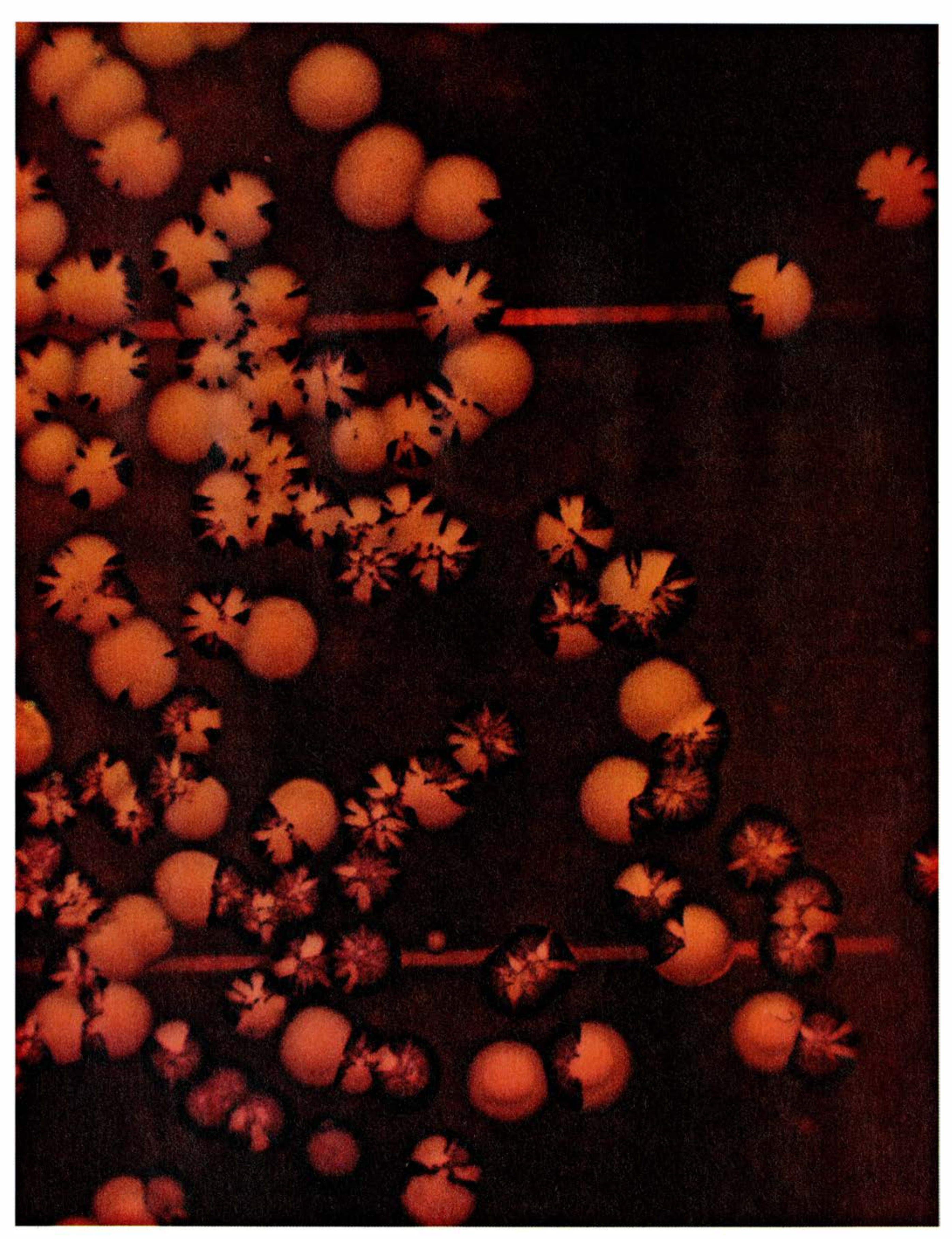
Uno de los hechos más notables de este período de cinco años ha sido el grave resurgimiento de la malaria, especialmente en el Asia sudoriental. El número de casos registrados se dobló, a la vez que muchos países reducían sus programas contra la malaria después de una campaña de erradicación que tuvo un éxito parcial. Se estima que en el Africa tropical mueren anualmente, al menos, medio millón de jóvenes menores de 14 años. La transmisión de la malaria sólo puede interrumpirse eliminando los vectores, parásitos del mosquito *Anopheles.* La ubicuidad del mosquito en muchas zonas del globo, el costo de los insecticidas y la creciente resistencia de los mosquitos a éstos multiplican las dificultades de la erradicación. Otra medida a tomar sería el uso generalizado de drogas antimaláricas; también aquí el problema son los costos, junto a la resistencia cada vez mayor del parásito a algunas drogas. Al igual que sucede con las vacunas, la obtención de nuevas drogas, de acuerdo con el informe de la OMS, requiere unos 10 años de investigación.

Además de la malaria, otras muchas enfermedades parasitarias, que en los países desarrollados no se recuerdan más que por sus extrañas denominaciones, forman parte de la vida diaria de muchos países subdesarrollados. La esquistosomiasis es endémica en 71 países; su agente causal, un gusano que parasita en los caracoles, infecta a unos 200 millones de personas; otros gusanos provocan la filariosis, que afecta al menos a 250 millones de personas, y la oncocercosis, una de las principales causas de ceguera en Africa.

La tuberculosis y seis enfermedades infantiles –la difteria, la tos ferina, el tétanos, el sarampión y la poliomielitis– pueden evitarse por inmunización, pero menos del 10 por ciento de los 80 millones de niños que nacen anualmente reciben una inmunización completa. Se supone que estas cinco enfermedades causan la muerte a unos cinco millones de niños menores de cinco años y ciegan, paralizan o incapacitan a otros cinco millones. En particular, ha habido un amenazador aumento en el número de casos de poliomielitis paralizante. Un objetivo principal de la OMS es lograr la inmunización de todos los niños frente a estas seis enfermedades en 1990.

En los países desarrollados, las enfermedades del corazón y de los vasos sanguíneos continúan siendo la principal causa de muerte entre los varones y la segunda o tercera entre las mujeres. Una buena noticia es que la tasa de mortalidad a causa de enfermedades cardiovasculares se ha estabilizado en la mayoría de países y ha descendido en algunos, especialmente en los Estados Unidos, por razones que aún no han podido aclararse. Su contrapartida es que las enfermedades cardiovasculares empiezan a adquirir cierta importancia en algunos países subdesarrollados. En éstos, la incidencia no se refleja principalmente en ataques cardíacos, como en los países desarrollados, sino en la lesión de las válvulas del corazón subsecuente a fiebres reumáticas, lo que a su vez complica las infecciones estreptocócicas. La fiebre reumática puede evitarse tratando la infección primaria con penicilina; el mantenimiento con penicilina de los niños o jóvenes con un historial de fiebres reumáticas los protege de posibles reinfecciones que provocarían o exacerbarían la lesión de las válvulas cardíacas. Por tanto, pueden evitarse los efectos lesivos, y a menudo fatales, del reumatismo cardíaco. La labor a realizar consiste en identificar las infecciones de estreptococos, administrar penicilina y asegurar un suministro mensual de penicilina a los jóvenes. En muchas partes del mundo, eso no es tarea fácil.

El cáncer, hasta hace poco un problema sanitario exclusivo de los países desarrollados, empieza a ser también una causa destacada de mortalidad en países subdesarrollados, como China, donde se está logrando controlar las enfermedades infecciosas y parasitarias. El informe de la OMS señala que la prevención real del cáncer no será, probablemente, simple; aun cuando se identifique un factor de riesgo de incidencia directa, su eliminación puede que suponga la alteración de hábitos fuertemente arraigados en la sociedad. "Una exposición excesiva a los rayos solares, ha podido descubrirse, provoca gran parte de los casos de cáncer de piel que se dan entre los caucasianos, pero desaconsejar la práctica de los baños de sol es una medida impopular." El factor de riesgo implicado con mayor claridad es el consumo de tabaco, que aumenta significativamente, además del riesgo de cáncer de pulmón, el riesgo de cáncer de laringe, de esófago y de vejiga (así como el riesgo de bronquitis, enfisema y de enfermedades coronarias); se estima que provoca más del 40 por ciento de las muertes por cáncer en varones, en Inglaterra. Otra reciente publicación de la OMS sugiere que el consumo de tabaco debiera considerarse pandémico, "pues el fumar es una forma no natural de conducta, que provoca una alteración del bienestar, y a menudo graves enfermedades, y se extiende por todo el mundo a través de un proceso de contaminación psicológica".



PRESENCIA DE PLASMIDOS en las bacterias y su comportamiento autónomo, según revela esta fotografía de colonias de *Staphylococcus aureus* que se desarrollan en un gradiente de temperaturas, de 30 grados (*abajo*) a 40 grados Celsius (*arriba*). La mayoría de las células sembradas inicialmente sobre el agar contenían un plásmido con un gen que codifica para una penicilinasa. Las células portadoras del plásmido y productoras de penicilinasa se tiñen de morado; las que no lo tienen, se tiñen de naranja y no producen penicilinasa. La replicación del plásmido, pero no la de las células, es sensible a la temperatura, de modo que cuanto mayor sea la temperatura, menos se replicará el plásmido y con mayor frecuencia aparecerán células que carecen de plásmido y de penicilinasa. Si una célula ha perdido el plásmido, sus descendientes se tiñen de naranja y las colonias tienen sectores de cada color.

Plásmidos

Estos elementos genéticos accesorios de las bacterias, conocidos sobre todo como portadores de resistencia a antibióticos y como vectores en ingeniería genética, son organismos subcelulares situados en el umbral mismo de la vida

Richard P. Novick

La célula es el único ambiente natural donde puede desarrollarse la intrincada química de la vida. Es, por ello, la unidad fundamental de organización biológica de los organismos unicelulares y pluricelulares. Representa también un nicho ecológico para diversos organismos subcelulares y submicroscópicos, tan apartados de los celulares en su estructura y dinámica reproductora que quizá debieran encuadrarse en un nuevo reino taxonómico.

Las más conocidas de estas curiosas formas de vida son los virus, partículas de elegante simetría compuestas por una molécula de ácido nucleico encerrada en una cápside de proteína. Pueden subsistir fuera de la célula, pero en ese entorno son inertes. Una vez insertadas en la célula, su ácido nucleico (ARN o ADN) reprograma, al servicio del virus, el aparato metabólico celular; el ácido nucleico se replica y se encapsida. El proceso finaliza con la destrucción de la célula y la liberación de una nueva cosecha de virus.

Hay una clase menos familiar de organismos subcelulares, los plásmidos, que se encuentran, en su mayoría, en bacterias. Los plásmidos tienen una organización más sencilla que los virus. Carecen de cápside proteínica; no poseen tampoco ninguna fase extracelular; se trata de moléculas circulares de ADN bicatenario que se multiplican independientemente en las células que los alojan y se heredan con regularidad al dividirse éstas. Tras su descubrimiento, hace unos 30 años, se creyó por algún tiempo que los plásmidos constituían una rareza. Ahora se sabe que son muy frecuentes. Se encuentran plásmidos en casi todas las especies de bacterias y probablemente en la mayoría de sus células. No son esenciales para el metabolismo usual de sus huéspedes, por lo que no es difícil aislar una variante bacteriana carente de plásmidos a partir de una estirpe que los tenga. La pérdida del plásmido es permanente: la célula no puede regenerar un nuevo plásmido, sólo puede adquirirlo de otra bacteria. La información contenida en un plásmido (o en cualquier otra molécula de ADN) sólo surge del largo y tortuoso proceso de la evolución.

Los plásmidos constituyen sólo una parte pequeña del genoma celular: en general, entre un 1 y un 2 o un 3 por ciento. Aun así, esa pequeña fracción de la información hereditaria determina rasgos genéticos accesorios, pero importantes, no codificados usualmente por el cromosoma de la bacteria. Sólo los plásmidos portan información para la conjugación ("apareamiento") entre bacterias. Tienen toda la responsabilidad de varias enfermedades en vegetales y animales, permiten que sus huéspedes se nutran de muchas sustancias complejas y les confieren resistencia a muchos tóxicos, antibióticos incluidos.

La importancia de estos rasgos genéticos, y en particular el interés clínico de la resistencia a antibióticos, atrajo la atención de los biólogos sobre los plásmidos; se comenzó entonces a comprender su papel en la genética bacteriana. En los últimos cinco años, se han investigado y manipulado intensamente los plásmidos, aunque por otra razón: se descubrió que eran los vehículos ideales para introducir en bacterias genes no bacterianos, lo que, mediante técnicas de clonación de moléculas, permitía la purificación y multiplicación de genes extraños. La clonación de ADN ha revolucionado ya el análisis genético y constituye la base de una nueva genética aplicada, que tiene grandes perspectivas médicas e industriales, pues logra que las bacterias fabriquen proteínas no bacterianas.

En una perspectiva más amplia, los plásmidos son también importantes porque han forzado a los biólogos a reconsiderar la naturaleza esencial de la vida. Los plásmidos están en el umbral mismo de la vida, entre lo inanimado y lo animado. ¿Son sólo trocitos de ADN, una parte integrante del genoma celular que aporta algo de información genética adicional? ¿O por el contrario son organismos subcelulares que evolucionan independientemente, más o menos como lo hacen los virus? Varias observaciones significativas apoyan con fuerza esta última interpretación.

Los plásmidos como organismos

Para empezar, cada plásmido controla autónomamente el número de sus copias que debe haber en la célula huésped. Además, la conjugación entre células bacterianas puede llevar al intercambio de plásmidos entre especies, e incluso géneros, diferentes, totalmente incapaces de intercambiar genes cromosómicos. Por último, este intercambio puede conducir a la transferencia, de una especie a otra competidora, de un plásmido portador de genes que permitan a la receptora sobrevivir a expensas de la donante. En otras palabras, los plásmidos han adquirido la capacidad de sobrevivir con independencia del destino de la especie que los aloja, algo inconcebible, dentro de la evolución por selección natural, en un mero componente del genoma de un organismo.

De estas observaciones se desprende que los plásmidos son verdaderos organismos independientes, miembros de una jerarquía de vida subcelular. Dentro de esta jerarquía, algunos virus de animales, como los de la poliomielitis o la fiebre amarilla, y bacteriófagos, o virus bacterianos, como el T4, pueden considerarse depredadores, pues matan siempre a las células que infectan. Los fagos atemperados y algunos virus de animales, que unas veces matan a sus huéspedes pero otras coexisten de forma estable en el interior de las

células sin causarles daño o beneficio aparente, se pueden considerar parásitos subcelulares. Los plásmidos serían endosimbiontes: residen establemente en la célula huésped y a menudo ésta se beneficia de un aporte de funciones genéticas adaptativas, el do ut des del estado simbiótico.

Considerar organismos a meras moléculas de ADN no encaja con el concepto convencional de la vida, que fija su base en la célula. Pero el rasgo común a todos los seres vivos es la replicación autónoma de su material genético y su evolución independiente. Por tanto, debiera considerarse organismo cualquier ácido nucleico que controle su propia replicación. El examen detallado de los procesos moleculares que determinan el ciclo vital de los plásmidos potenciará sin duda nuestra comprensión de la naturaleza de la vida e iluminará la frontera entre lo animado y lo inanimado.

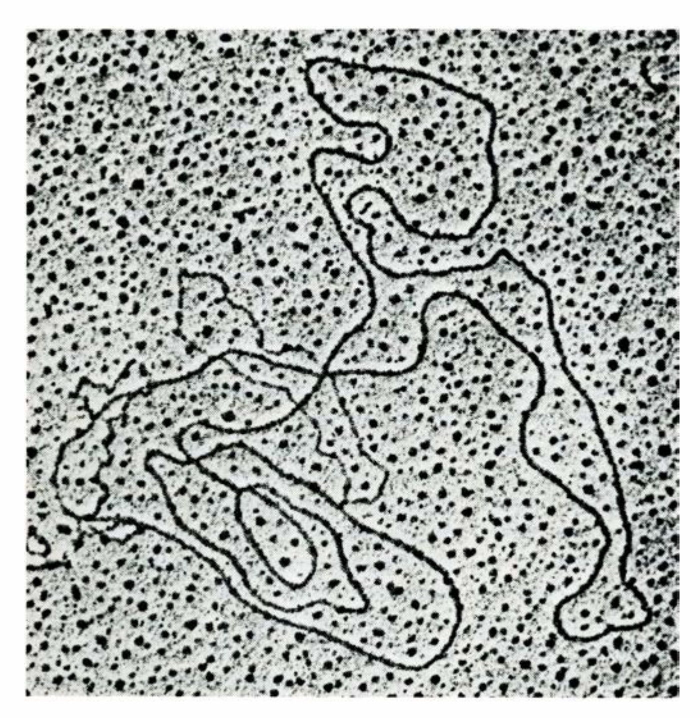
El descubrimiento de los plásmidos

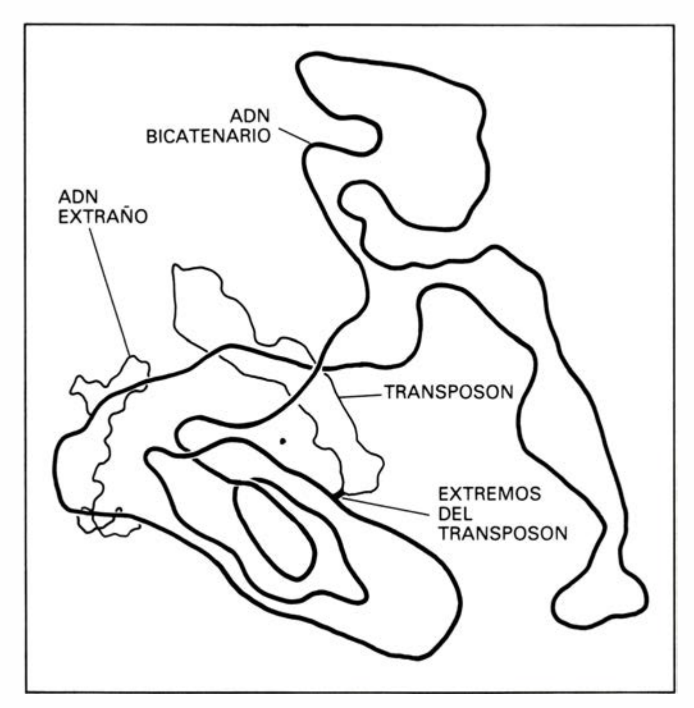
A principios de la década de 1950, poco después de que Joshua Lederberg descubriera la conjugación en la bacteria *Escherichia coli,* quedó claro que en esta bacteria había dos tipos "sexuales" determinados genéticamente y que la información genética se transfería físicamente del donante, o "macho", al receptor, o "hembra". El factor F (de "fertilidad"), responsable de la masculinidad, se transfería mucho más a menudo que cualquier otro rasgo genético. Además, el factor F se transfería independientemente de los demás genes conocidos del donante: no estaba ligado a ninguno de ellos. De tratarse de una célula eucariota (una célula dotada de núcleo y con su ADN organizado en cromosomas visibles al microscopio óptico) este resultado hubiera significado que el carácter de "masculinidad" estaba codificado por un gen extracromosómico. Lederberg observó que el factor F presentaba cierta semejanza con los elementos genéticos extracromosómicos del citoplasma de los organismos superiores; en 1952 acuñó el término plásmido para referirse a todos los sistemas genéticos extracromosómicos de este tipo. En aquel entonces no se conocía bien la organización estructural del genoma de las bacterias; ello explica que no se supiera a ciencia cierta la significación precisa de la ausencia de ligamiento.

Pero el nudo del problema se deshizo pronto, gracias a la subsiguiente investigación del factor F. En 1953, William Hayes, a la sazón en el Hospital Hammersmith de Londres, descubrió que en ciertas circunstancias el factor F aparecía ligado a marcadores normales, lo que provocaba la transferencia ordenada de éstos durante la conjugación. Dedujo que la mayoría de los genes de *E. coli* residían en una sola estructura continua que se comportaba en términos genéticos formales como un círculo: la analogía bacteriana del cromosoma eucariótico. El factor F se unía físicamente a esta estructura en un lugar específico; durante la conjugación, el círculo se abría por el lugar de inserción de F y se transfería como una estructura lineal al receptor. También se observó que el factor F inserto podía liberarse en alguna ocasión, arrastrando a veces consigo un segmento del cromosoma bacteriano contiguo a su lugar de inserción. El factor F podía considerarse un elemento específicamente extracromosómico, que a veces se integraba en el cromosoma.

En el Instituto Pasteur de París, François Jacob y Elie L. Wollman apreciaron cierta semejanza entre el comportamiento del factor F, el del fago atemperado lambda y el de otro plásmido, ColE1 (responsable de una colicina, proteína que mata a *E. coli*). Acuñaron un nuevo término, episoma, para los elementos genéticos que pueden replicarse en dos estados alternativos: unidos al cromosoma del huésped o libres en su citoplasma.

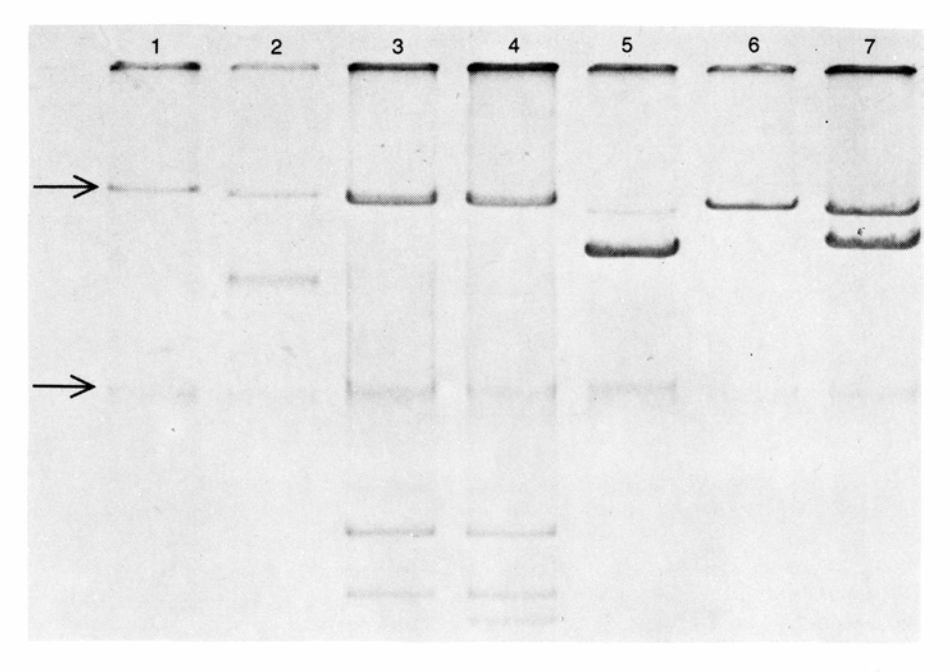
Por un tiempo, el fago lambda, el factor F, ColE1 y uno o dos elementos más fueron los únicos episomas conocidos. Pero en 1959 tuvo lugar en Japón un descubrimiento notable mientras se examinaba el caso de pacientes de disentería bacteriana que no respondían al tratamiento con antibióticos usualmente eficaces. La bacteria cau-





HETERODUPLEX PLASMIDICO según una electromicrografía obtenida por Marjorie M. Schwesinger en el laboratorio del autor. Se desnaturalizaron y separaron las cadenas de ADN de los plásmidos pI258 y pI6187, idénticos salvo en que el primero porta un transposón que determina resistencia a la eritromicina; luego, se dejaron asociar las cadenas separadas. El heterodúplex tiene una cadena de cada uno de los plásmidos. Las dos cadenas se emparejan perfectamente, a excepción de la secuencia del transposón. Como indica el esquema, las dos cadenas forman una molécula bicatenaria con un lazo unicatenario, el transposón, que cuelga del corto segmento bicatenario formado por las "terminaciones de secuencia repetida e invertida".

sante del trastorno, *Shigella dysenteriae,* portaba genes que le conferían resistencia simultánea a varios antibióticos; estos genes, además, se transferían a otras bacterias intestinales de forma parecida a la del factor F. Estos factores, llamados R (de "resistencia"), compartían varias propiedades con el factor F, entre ellas la capacidad de promover su propia transferencia intercelular por conjugación; después se ha sabido que algunos de ellos coinciden con F en segmentos de la secuencia de ADN. Mientras que el factor F se había considerado una entidad algo especial, muy interesante desde el punto de vista genético pero no necesariamente muy difundido, el descubrimiento de los factores R hizo ver que los elementos genéticos extracromosómicos eran componentes importantes del microcosmos bacteriano, con interés clínico amén de científico.



ELECTROFORESIS EN GEL: ensayo que revela la presencia de plásmidos en bacterias. Daniel A. Portnoy y Stanley Falkow, de la Universidad de Washington en Seattle, prepararon extractos de siete estirpes de la bacteria *Escherichia coli,* eliminaron la mayoría del ADN cromosómico y colocaron muestras en pocillos practicados en un extremo de una placa de gel de agarosa. Una corriente eléctrica hizo que las moléculas de ADN, cargadas negativamente, emigraran hacia el electrodo positivo a una velocidad inversamente proporcional al logaritmo de su tamaño. El ADN, marcado con un colorante fluorescente, aparece en forma de bandas. La banda superior representa material retenido en origen. La siguiente banda (*flecha superior*) representa un plásmido responsable de la diarrea de los viajeros, común a todas las muestras excepto la número 5, que es una cepa no patógena. Las demás bandas nítidas corresponden a plásmidos diversos, algunos de los cuales determinan resistencia a antibióticos. Las bandas más difusas (*flecha inferior*) están formadas por fragmentos de ADN cromosómico residual.

Poco después del descubrimiento de los factores R en las enterobacterias encontramos factores similares en los estafilococos. Contenían éstos un gen que codificaba para un poderosísimo enzima, una beta-lactamasa de la penicilina (o penicilinasa), que destruía la penicilina, y eran responsables de la resistencia a la penicilina que estaba dificultando el tratamiento de infecciones hospitalarias, en muchas partes del mundo, en los primeros años del decenio de los sesenta. Los factores R de los estafilococos parecían incapaces de promover su propia transferencia por conjugación. Se transferían pasivamente por transducción, proceso en el que un fragmento de ADN bacteriano se encapsida en una partícula de fago, en lugar del ADN propio del fago, y se introduce en otra célula infectada por el fago. Una proporción importante de los factores R de las enterobacterias tampoco puede promover su propia transferencia conjugativa y depende de la transducción o de la movilización por otro plásmido conjugativo que se encuentre en la misma célula.

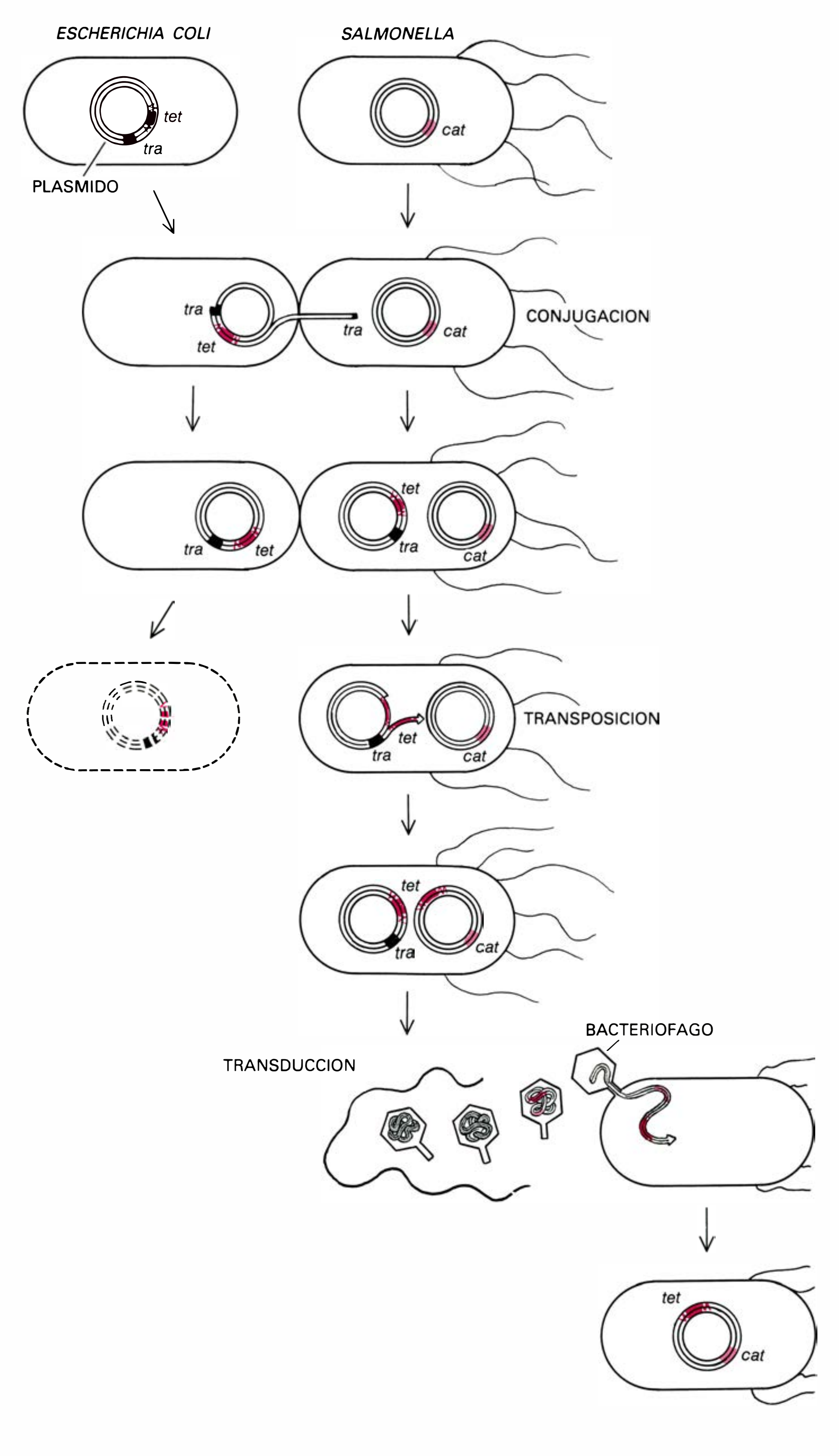
Mediados ya los sesenta, se sabía que la mayoría de los factores R de los estafilococos y las enterobacterias (y también ColE1) se diferenciaban del factor F y de lambda en que eran siempre extracromosómicos y no se integraban reversiblemente en el cromosoma; por tanto, no satisfacían la definición restrictiva de episoma. Me pareció que la exigencia de integración reversible impedía agrupar muchos sistemas genéticos extracromosómicos claramente relacionados; por ello, en 1963, sugerí que quizá fuera más apropiado el término, más general, propuesto por Lederberg; otros investigadores estuvieron de acuerdo y, poco a poco, a lo largo de los siguientes 15 años, los elementos genéticos extracromosómicos móviles de las bacterias vinieron a llamarse plásmidos en vez de episomas.

Variedad de los plásmidos

Poco a poco, fue haciéndose también evidente que los plásmidos eran responsables de una sorprendente riqueza de actividades biológicas en las bacterias. La valoración del papel desempeñado por los plásmidos (y por otros sistemas genéticos extracromosómicos descubiertos más recientemente) ha dado lugar a un cambio fundamental en la interpretación que los biólogos dan a los procesos genéticos. La explicación tradicional sostenía que la constitución genética de una especie variaba poco de una célula a otra y que permanecía constante durante mucho tiempo. Se sabe ahora que una proporción significativa de los rasgos genéticos, no sólo de las bacterias sino también de los organismos superiores, son variables (presentes en algunas células o estirpes y ausentes en otras), lábiles (se adquieren o pierden con facilidad) y móviles (transferibles entre células o transponibles de un lugar a otro de una célula), todo ello debido a que estos rasgos están asociados con plásmidos y otros sistemas genéticos atípicos.

En 1964, Mark Richmond y Eric Johnston, del National Institute for Medical Research de Inglaterra, demostraron que los estafilococos responsables de una serie de infecciones de las suturas quirúrgicas llevaban un plásmido que les permitía no sólo destruir la penicilina, sino también crecer en el desinfectante mercuriado usado para esterilizar las suturas. En mi laboratorio del Public Health Research Institute de la ciudad de Nueva York, Christine Roth y yo observamos que nuestros plásmidos de estafilococos portaban genes que conferían resistencia a la penicilina, a compuestos de mercurio y a varios compuestos de otros metales pesados letales para los estafilococos, como las sales de cadmio, bismuto, plomo y antimonio y los iones arseniato y arsenito. Desde entonces, se han hallado diversos genes responsables de resistencia a metales pesados en algunos plásmidos R de *E. coli.*

Muchos otros tipos de genes especializados se encuentran en plásmidos y no en el cromosoma bacteriano; nos referimos, entre otros, a los genes responsables de ciertas enfermedades



GENES DE RESISTENCIA a antibióticos. Se acumulan en los plásmidos y se transfieren de una bacteria a otra por varios mecanismos, que a veces permiten a un plásmido sobrevivir a la muerte de su huésped. Las dos bacterias representadas (*arriba*) son una célula de *Escherichia coli*, que contiene un plásmido con genes para su transmisión por conjugación (*tra*) y resistencia a la tetraciclina (*tet*), y una célula de *Salmonella*, con un plásmido que determina un gen de resistencia al cloranfenicol (*cat*). Durante la conjugación, el plásmido que lleva *tet* pasa a *Salmonella* y la hace resistente a la tetraciclina, además de al cloranfenicol. En presencia de ambos antibióticos, *E. coli* muere, pero su plásmido sobrevive en el nuevo huésped. El gen *tet* se encuentra en un transposón que más tarde pasa de un plásmido al otro, que reúne así los genes de resistencia a ambos antibióticos. El plásmido se puede transferir también por transducción: un virus bacteriano infecta a *Salmonella* y se reproduce, matándola; una de las partículas del bacteriófago incorpora el plásmido en vez del ADN del virus y lo transfiere a una nueva célula.

humanas tales como la diarrea de los viajeros y el impétigo por estafilococos, los encargados de la fermentación bacteriana de la leche en la producción de queso y de gran parte de la versatilidad metabólica de las pseudomonas. Las pseudomonas son bacterias del suelo capaces de degradar diversos hidrocarburos complejos, cuyo concurso podría aprovecharse en la limpieza de las manchas de petróleo o en la conversión de éste en proteína comestible. Los plásmidos producen una curiosa vacuola gaseosa en las bacterias halófilas; también se cree que determinan, en *Bacillus thuringiensis,* la síntesis de un insecticida muy empleado en la lucha contra lepidópteros como la mariposa de la col.

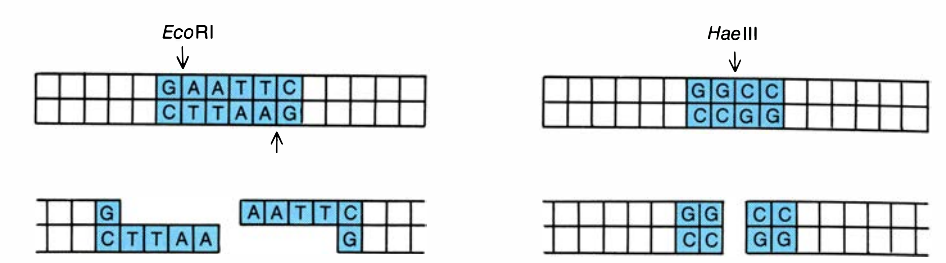
Aparte de la naturaleza algo especial de muchas de sus funciones, los genes de los plásmidos no se diferencian básicamente de los demás genes. La mayoría determina la estructura de proteínas enzimáticas, sintetizadas por la misma maquinaria metabólica celular que las proteínas restantes.

Resistencia a antibióticos

El más estudiado de los efectos subsiguientes de la presencia de plásmidos es la resistencia a antibióticos y a otras sustancias tóxicas para las bacterias. Fue ésta la primera propiedad atribuida a los plásmidos, que, además, representa un grave problema médico. La resistencia a una determinada substancia proporciona caracteres selectivos de gran aplicación en experimentos de genética bacteriana y clonación molecular; su investigación ilumina los orígenes de los plásmidos y el mecanismo mediante el cual logran ir acumulando diferentes genes.

La evolución ha desarrollado por lo menos tres estrategias diferentes por las que los genes de los plásmidos determinan resistencia a substancias tóxicas, antibióticos incluidos. La más sencilla es la simple destrucción del antibiótico; tal ocurre con la resistencia a la penicilina, cloramfenicol, aminoglicósidos (como la estreptomicina) y compuestos de mercurio. Una segunda estrategia consiste en impedir el acceso del tóxico a su objetivo celular. Los antibióticos macrólidos (como la eritromicina) se unen a los orgánulos celulares llamados ribosomas y detienen la síntesis de proteínas; Bernard Weisblum y sus colegas, de la Facultad de Medicina de la Universidad de Wisconsin, han demostrado que los genes de resistencia transportados por los plásmidos dan lugar a la adición de un

grupo metilo al ARN ribosómico, impidiendo así la unión del antibiótico. Los plásmidos pueden también determinar proteínas que modifiquen algún sistema específico de transporte de la membrana celular; según parece, así se impide la entrada de la tetraciclina en la célula. Simon D. Silver, de la Universidad de Washington, ha demostrado que la resistencia de los estafilococos al ion arseniato se debe a un proceso activo de bombeo, mediado de alguna manera por genes plasmídicos; los iones entran en las células sin dificultad, pero salen expulsados a la misma velocidad con la que entran.



ENDONUCLEASAS DE RESTRICCION O RESTRICTASAS. Por tales se entiende aquellos enzimas que cortan el ADN en lugares determinados por la secuencia de las bases que lo forman: *A, G, T* y *C*. Cada enzima (el esquema ilustra dos distintos) reconoce cierta secuencia palindrómica de bases y corta cada cadena del ADN en un lugar específico de esa secuencia. Si los cortes no están exactamente opuestos, como en el caso de *Eco*RI (*izquierda*), los extremos de los fragmentos producidos son unicatenarios y complementarios. Tales "extremos cohesivos" les permiten unirse a otros fragmentos producidos por el mismo enzima; esta propiedad puede aprovecharse para generar nuevas combinaciones de fragmentos.

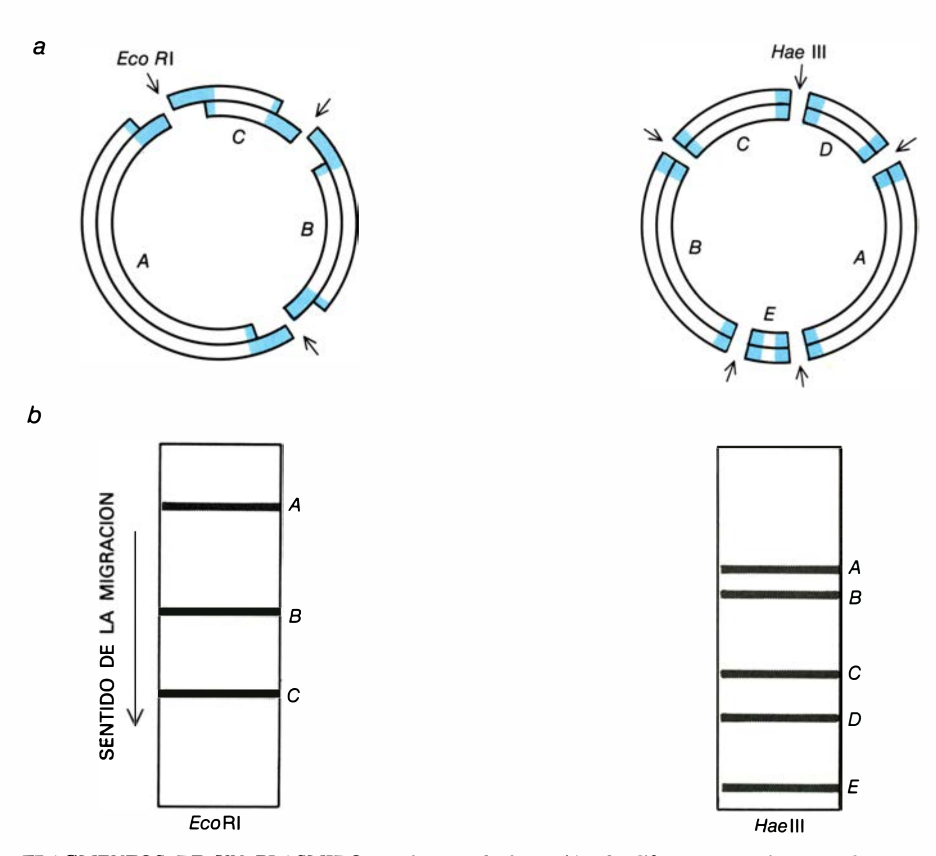
La tercera estrategia, creación de un proceso alternativo al proceso inhibido por el agente antibacteriano, se encuentra en el caso de la resistencia a las sulfamidas y a la trimetoprima, que inhiben los enzimas bacterianos necesarios para la síntesis de una vitamina esencial, el ácido fólico. (Los humanos, a diferencia de las bacterias, somos incapaces de sintetizar la vitamina y la tomamos de nuestros alimentos, aunque no en cantidad suficiente como para mantener a las bacterias que nos infectan.) La resistencia a estas drogas, determinada por los plásmidos, comporta la sustitución del enzima sensible a la droga por un enzima que, aunque tenga la misma función, sea insensible a la inhibición.

Desde luego, estas estrategias genéticas no han surgido de repente; deben haberse originado en el lento curso de la evolución y su antigüedad supera en millones de años a la aplicación clínica de los antibióticos. De hecho, se encuentran plásmidos R completamente desarrollados en bacterias aisladas y almacenadas mucho antes de la era de los antibióticos o aisladas recientemente en lugares remotos, donde probablemente no han estado expuestas a antibióticos.

Debo hacer constar que los plásmidos no son la única fuente de resistencia bacteriana a antibióticos. Este carácter puede obedecer también a mutaciones en genes cromosómicos. Durante varios años se creyó erróneamente que las mutaciones cromosómicas eran responsables de la resistencia clínica a los antibióticos. La confusión surgió de experimentos de laboratorio en los que se trataban con antibióticos cultivos puros de bacterias. Si se había dado una mutación espontánea en un cromosoma bacteriano que permitiera fortuitamente a la célula portadora resistir a la droga, esa célula proliferaba y las demás morían. Pero este mecanismo no lograba explicar la resistencia natural simultánea a varios antibióticos observada en bacterias presentes en una persona tratada con un solo antibiótico. Más aún, los mutantes resistentes aislados en experimentos de laboratorio solían crecer mal en ausencia del antibiótico empleado para seleccionarlos. Las mutaciones cromosómicas de resistencia modifican el componente celular sensible al antibiótico, por ejemplo el ribosoma o la pared celular, reduciendo su eficacia general; los mutantes resistentes son engendros evolutivos que desaparecen rápidamente en condiciones naturales. Por contra, las resistencias determinadas por plásmidos han evolucionado con el tiempo hasta alcanzar una especificidad exquisita: protegen el objetivo celular del antibiótico sin reducir significativamente la adaptatividad general de la célula. Ahí radica la clave de su éxito.

Evolución de los plásmidos

¿Cómo y por qué aparecieron los genes de resistencia de los plásmidos? Julian E. Davies, de la Universidad de Wisconsin, notando que los antibióticos eran producidos por bacterias y hongos, sugirió que fueron los mismos microbios quienes desarrollaron genes



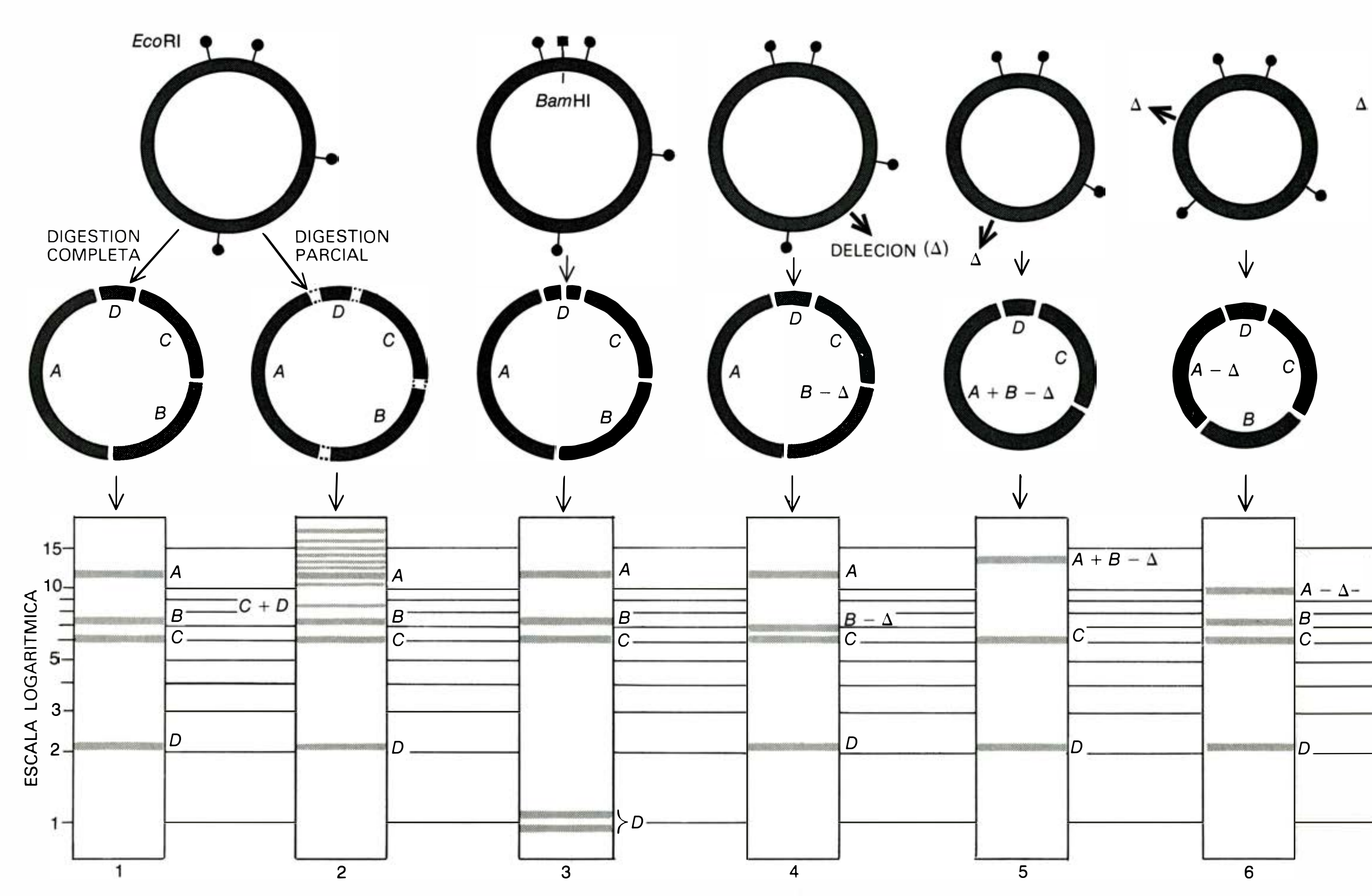
FRAGMENTOS DE UN PLASMIDO resultantes de la acción de diferentes restrictasas sobre sus respectivas secuencias sensibles (*a*). La electroforesis en gel permite determinar la longitud de los fragmentos y, por tanto, la situación de los puntos de corte. Tras migración por el gel y tinción del ADN, aparecen bandas (*b*), integradas cada una por fragmentos del mismo tamaño. Su distancia del origen indica el tamaño de los fragmentos que la componen: cuanto mayor es la distancia, menor la molécula.

de resistencia para protegerse de sus propios productos. El y sus colaboradores demostraron que *Bacillus circulans,* productor de la butirosina, un antibiótico aminoglicósido, sintetizaba un enzima que inactivaba la butirosina y otros aminoglicósidos, como la neomicina. Se puede transferir el gen de este enzima a *E. coli* por técnicas de ADN recombinante, con lo que *E. coli* se vuelve resistente a la neomicina.

A mi entender existe otra posibilidad. Los antibióticos son fabricados por organismos del suelo. La presencia en el suelo de antibióticos y de trazas de metales pesados tóxicos, aun en pequeñas concentraciones, pudo haber provocado la evolución de genes de resistencia en las bacterias sensibles presentes en los suelos, ya fuera directamente como genes de plásmidos o como genes cromosómicos transferidos más tarde a plásmidos. En cualquier caso, los genes habrían pasado después a bacterias no edáficas causantes de infecciones en los animales.

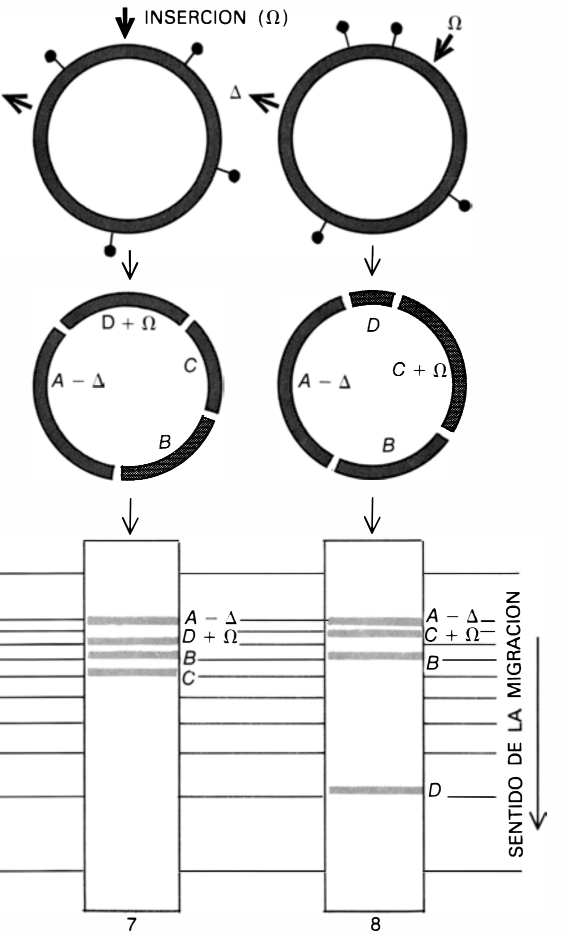
Apoya esta idea el hecho de que, aunque la penicilina la fabriquen los hongos (que son eucariotas, no bacterias), muchas bacterias no patógenas del suelo fabrican una penicilinasa parecida a la que hace resistentes a la penicilina a las bacterias patógenas de los animales. Se han encontrado plásmidos que determinan penicilinasa en cepas de estafilococos aisladas antes del descubrimiento de la penicilina. Y se han descubierto plásmidos que determinan resistencia a la tetraciclina en varias especies de bacterias del suelo. En mi laboratorio, June Polak ha observado la transmisión de uno de estos plásmidos a *Staphylococcus aureus,* que se vuelve así resistente a la tetraciclina; el plásmido es casi idéntico a otro nativo de los estafilococos que determina resistencia a la kanamicina. Esta hipótesis implica que la aplicación masiva de antibióticos en medicina y agricultura ha alterado un ecosistema muy antiguo, en el que se había alcanzado un equilibrio entre productores de antibióticos y organismos dotados de genes de resistencia.

Los primeros plásmidos R que se descubrieron solían determinar resistencia a uno o dos inhibidores. No es raro encontrar ahora plásmidos que confieren resistencia hasta a diez antibióticos. Una implicación clínica inquietante de esta acumulación de genes de resistencia es que el tratamiento de un paciente con un solo antibiótico puede seleccionar un organismo resistente a muchos otros. El análisis de este fenómeno ha revelado otra propiedad sorprendente de los plásmidos y los sistemas genéticos relacionados: la transposición de genes.



PARA OBTENER EL MAPA del plásmido pI258 se corta el plásmido y sus mutantes con restrictasas y se analizan los fragmentos por electroforesis en gel. En la fila superior se muestran los plásmidos y sus lugares de corte; en la siguiente, los productos del fraccionamiento. Bajo cada grupo de fragmentos aparece el esquema de un gel que nos da el tamaño de cada fragmento. El gel número 1 muestra que el enzima *Eco*RI corta el plásmido en cuatro fragmentos, *A, B, C* y *D,* que tienen respectivamente 13, siete, seis y dos kilobases (miles de pares de bases). El orden de los fragmentos en el círculo se determina por "digestión parcial", es decir, de forma que no todas las moléculas de plásmido queden totalmente seccionadas. La presencia en el gel número 2 de fragmentos que suman *C + D,* pero no *B + D,* establece la secuencia *ABCD.* Para construir un mapa más detallado se obtienen fragmentos más pequeños tratando la mezcla de trozos con otro enzima. El gel número 3, por ejemplo, muestra que el enzima *Bam*HI corta a *D* en dos, ubicando así el único lugar de rotura de *Bam*HI del plásmido. El plásmido del gel número 4 ha sufrido una deleción que altera la secuencia responsable de la penicilinasa; el gen de la penicilinasa debe estar, al menos en parte, en el fragmento *B,* porque la banda *B* es más pequeña (va más lejos). Otra deleción (gel número 5), que elimina la resistencia a los iones arseniato y arsenito, favorece la unión de los fragmentos *A* y *B,* confirmando su contigüidad y localizando los genes de la resistencia en o cerca del límite de ambos. Un transposón que confiere resistencia a la eritromicina se ha perdido en el

Por analogía con el comportamiento del factor F, se supuso inicialmente que los plásmidos R habían adquirido mutaciones cromosómicas de resistencia tras haberse integrado temporalmente en el cromosoma y haberse liberado llevándose un segmento de ADN cromosómico. Pero no se lograron encontrar en los plásmidos secuencias de ADN propias de los cromosomas de sus huéspedes, como cabría esperar si los plásmidos hubieran adquirido cantidades significativas de ADN cromosómico. Además, si los plásmidos pudieran en verdad adquirir rutinariamente genes cromosómicos, ¿cómo parecían haber adquirido sólo genes que determinaban resistencia? Incluso bajo la actuación de fuerzas selectivas, deberían haber arrancado también secuencias contiguas a los genes de resistencia. En realidad, según se supo después, muchos plásmidos R no se insertan en el cromosoma de la forma en que lo hacen los factores F.



plásmido del gel número 6; tenía que haber estado en *A* porque la banda *A* viaja ahora más deprisa. Las inserciones secundarias del transposón en el fragmento *D* (gel número 7) o en el fragmento *C* (gel número 8) abolen la resistencia a desinfectantes mercuriados. El gen responsable de esta resistencia debe estar, por tanto, parte en el fragmento *D* y parte en el *C*. Sólo el conocimiento detallado de la secuencia del plásmido, que resulte de la aplicación de estas técnicas, permitirá su plena comprensión como entidad genética.

La verdadera razón de la resistencia múltiple a antibióticos no se conoció hasta 1974; los primeros en dar con ella fueron Alan E. Jacob y Robert W. Hedges, de la Real Facultad de Medicina para Postgraduados de la Universidad de Londres; poco después fue confirmada por no menos de otros cinco grupos de investigadores. Descubrieron que los genes de resistencia en cuestión se encontraban en unos elementos genéticos nuevos, con capacidad para transferirse físicamente de un lugar a otro del material genético de una célula, sin necesidad de secuencias complementarias de ADN y sin utilizar los enzimas implicados en los mecanismos de recombinación usuales. Estos nuevos elementos genéticos, los transposones, eran responsables no sólo de la acumulación de genes de resistencia en los plásmidos, sino también de grandes reorganizaciones del material genético en bacterias y, al parecer, también en ecuariotas [véase "Elementos genéticos transponibles", por Stanley N. Cohen y James A. Shapiro; INVESTIGACION Y CIENCIA; abril, 1980]. Probablemente, el predominio de los genes de resistencia en los transposones descubiertos hasta ahora resultará tener la misma base que la abundancia de los propios plásmidos R: el cambio evolutivo radical impuesto por la diseminación masiva de antibióticos.

Estructura de los plásmidos

La organización de los genes de los plásmidos parece hacer compatible la estabilidad hereditaria con una gran flexibilidad estructural. A diferencia de los virus, los plásmidos pueden adquirir nuevos genes y reorganizar extensamente los antiguos para mantener una información genética apropiada a las necesidades de su huésped del momento, sin comprometer la eficacia de su propia replicación. Para apreciar en su medida la estrategia genética de los plásmidos hay que saber algo del análisis experimental que la ha revelado.

Las técnicas que han revelado la estructura y función de los sistemas genéticos de los animales, las plantas e incluso los virus no dan mucho resultado con los plásmidos. Esas técnicas se basan en el estudio de la transmisión de mutaciones génicas en cruzamientos experimentales. Es difícil aislar mutaciones que afecten a la replicación de los plásmidos y el análisis por recombinación genética desalienta por su complejidad. Por el contrario, los plásmidos pueden estudiarse adecuadamente con algunas técnicas recientes de la genética molecular, en particular, la aplicación de la electroforesis en gel al análisis del ADN y la disección y clonación de secuencias de ADN con ciertos enzimas bacterianos llamados endonucleasas de restricción, o restrictasas.

Un plásmido, como la mayoría de las demás moléculas de ADN, es una doble hélice, en la que cada cadena constituye una secuencia lineal de cuatro bases nucleotídicas: adenina (*A*), guanina (*G*), timina (*T*) y citosina (*C*). Las secuencias de bases de las dos cadenas son complementarias: *A* siempre se empareja con *T* y *G* con *C*. Cada restrictasa reconoce cierta secuencia breve de bases dondequiera que se encuentre y rompe el ADN en un sitio determinado de esa secuencia. Dado un gran número de moléculas idénticas de ADN circular, por ejemplo una muestra de ADN purificado de un plásmido, una restrictasa cortará todas las moléculas por los mismos sitios, produciendo tantos fragmentos cuantos lugares de rotura haya en la molécula. La electroforesis en gel permite separar fácilmente unos fragmentos de otros según su tamaño (o longitud de la molécula). Se coloca el ADN fragmentado en un pocillo rectangular excavado en un extremo de una placa de gel y se aplica una corriente eléctrica; ello provoca que las moléculas de ADN emigren a través del gel hacia el electrodo positivo a una velocidad inversamente proporcional al logaritmo de su longitud molecular. Los fragmentos idénticos forman una estrecha banda horizontal, cuya posición puede observarse empapando el gel con un colorante fluorescente, como el bromuro de etidio (que se une al ADN), y fotografiándolo bajo luz ultravioleta.

La disección de un plásmido con varias restrictasas, a la vez o por separado, proporciona varios conjuntos de fragmentos que pueden ordenarse en un mapa de fragmentación. Los genes de los plásmidos pueden ubicarse en este mapa de varias maneras; por ejemplo, analizando plásmidos que tengan mutaciones estructurales identificables, como deleciones o inserciones. Aclarará mejor el procedimiento la consideración de los primeros pasos del análisis del plásmido pI258, un factor R estafilocócico, que se extiende en una longitud total de unos 28.000 pares de bases y contiene genes de resistencia a la eritromicina, penicilina y a varios compuestos metálicos tóxicos.

Se trata primero el plásmido con la restrictasa *Eco*RI, obteniendo cuatro fragmentos de 13, siete, seis y dos kilobases (miles de pares de bases), según refleja el gel número 1 [*véanse las ilustraciones de las dos páginas precedentes*]. La posición real de estos fragmentos en el círculo se determina por digestión parcial (gel número 2). Mediante otros ensayos puede determinarse el punto de disección de otros enzimas de restricción. En el gel número 3 se aprecia uno de ellos. Los demás geles muestran los fragmentos obtenidos al tratar plásmidos mutantes, de los que se han eliminado fragmentos que afectan a rasgos conocidos o en los que se ha insertado un transposón. El examen de un plásmido con una deleción que afecte a un determinado gen nos permite precisar qué fragmento es el acortado y, de esta forma, localizar el gen afectado. La inserción de un transposón alarga el fragmento e impide la función del gen en que se insertó, señalando también la situación del gen inactivado.

Examinando sólo unas pocas reordenaciones del ADN del plásmido se puede aprender mucho sobre su estructura y organización genética. Combinando más experimentos de electroforesis con la determinación directa de la secuencia completa de bases del plásmido (que se ha conseguido sólo en plásmidos mucho más pequeños) se espera obtener la detallada información necesaria para la plena comprensión del plásmido como entidad genética.

Autonomía de los plásmidos

Los plásmidos representan el paroxismo de la simbiosis, porque dependen del huésped para todas las funciones vitales esenciales, salvo las "funciones autonómicas". Son éstas las funciones más íntimamente relacionadas con su aspiración a la categoría de organismo: aquellas que sirven para definir su individualidad. La identificación y el análisis de estas funciones revela, por tanto, los componentes esenciales mínimos de la vida. En otras palabras, todo plásmido debe contener en su genoma una solución al problema de cómo puede heredarse con regularidad un sistema genético no esencial y cómo puede éste controlar su propia replicación y la distribución de sus copias para mantener una relación constante con su huésped. ¿Cuáles son estas soluciones?

La base teórica necesaria para abordar este problema fue desarrollada en 1963 por Jacob, Sydney Brenner y François Cuzin. Propusieron que en todo "replicón", o unidad de replicación, ya fuera un cromosoma bacteriano, un segmento de cromosoma eucariótico, un virus, un plásmido o cualquier otra forma de ácido nucleico, la replicación comenzaba en un punto específico (el origen), avanzando sucesiva y linealmente hasta un punto fijo de terminación; una sustancia "iniciadora", codificada por el propio replicón, actuaría sobre el origen para poner en marcha el proceso de la replicación. Propusieron que, en las bacterias, todos los replicones autónomos debían estar unidos permanentemente a la superficie interior de la membrana celular en un lugar próximo al plano ecuatorial de la célula; esta unión sería esencial para la replicación y para el reparto de las copias a las células hijas. En un momento determinado del ciclo celular aparecería un nuevo lugar de unión para cada replicación al otro lado del plano ecuatorial, lo que desencadenaría un ciclo de replicación, durante el cual, una molécula de ADN recién sintetizada se uniría al nuevo punto de anclaje. El crecimiento de un septo de división en el plano ecuatorial distribuiría una molécula de ADN a cada célula hija. Esta brillante hipótesis, muy avanzada para su época y formulada sin una excesiva base experimental, ha guiado virtualmente todos los estudios posteriores sobre la regulación de la replicación bacteriana y sobre la base genética de la autonomía de los plásmidos.

Predecía la teoría que unas cuantas funciones genéticas, tan sólo una pequeña fracción de las existentes en cualquier plásmido, bastarían para asegurar la autonomía de éste. El aislamiento de plásmidos mutantes que a pesar de haber perdido gran parte de su ADN seguían siendo viables lo confirmó. Estudiando las deleciones del plásmido pI258 de estafilococos observamos que podía perder hasta dos tercios de su dotación génica sin menoscabo de su autonomía. Ningún otro tipo de genoma puede soportar deleciones de tal magnitud. Sí había, no obstante, una pequeña región "prohibida", presente en todos los plásmidos viables y que resultó contener todas las funciones autonómicas; la deleción de cualquier parte de esta región destruía la viabilidad del plásmido.

Se han obtenido resultados parecidos con muchos otros plásmidos y ha quedado bien establecido que los genes responsables de la autonomía de los plásmidos están agrupados en lo que Ken-ichi Matsubara, de la Universidad de Osaka, ha denominado "bloque replicador". En 1966, cuando concluí mis experimentos, las técnicas moleculares no habían avanzado lo bastante para la determinación exacta del tamaño del bloque replicador. Se sabe ahora que no supera los 3000 pares de bases: la décima parte del plásmido pI258.

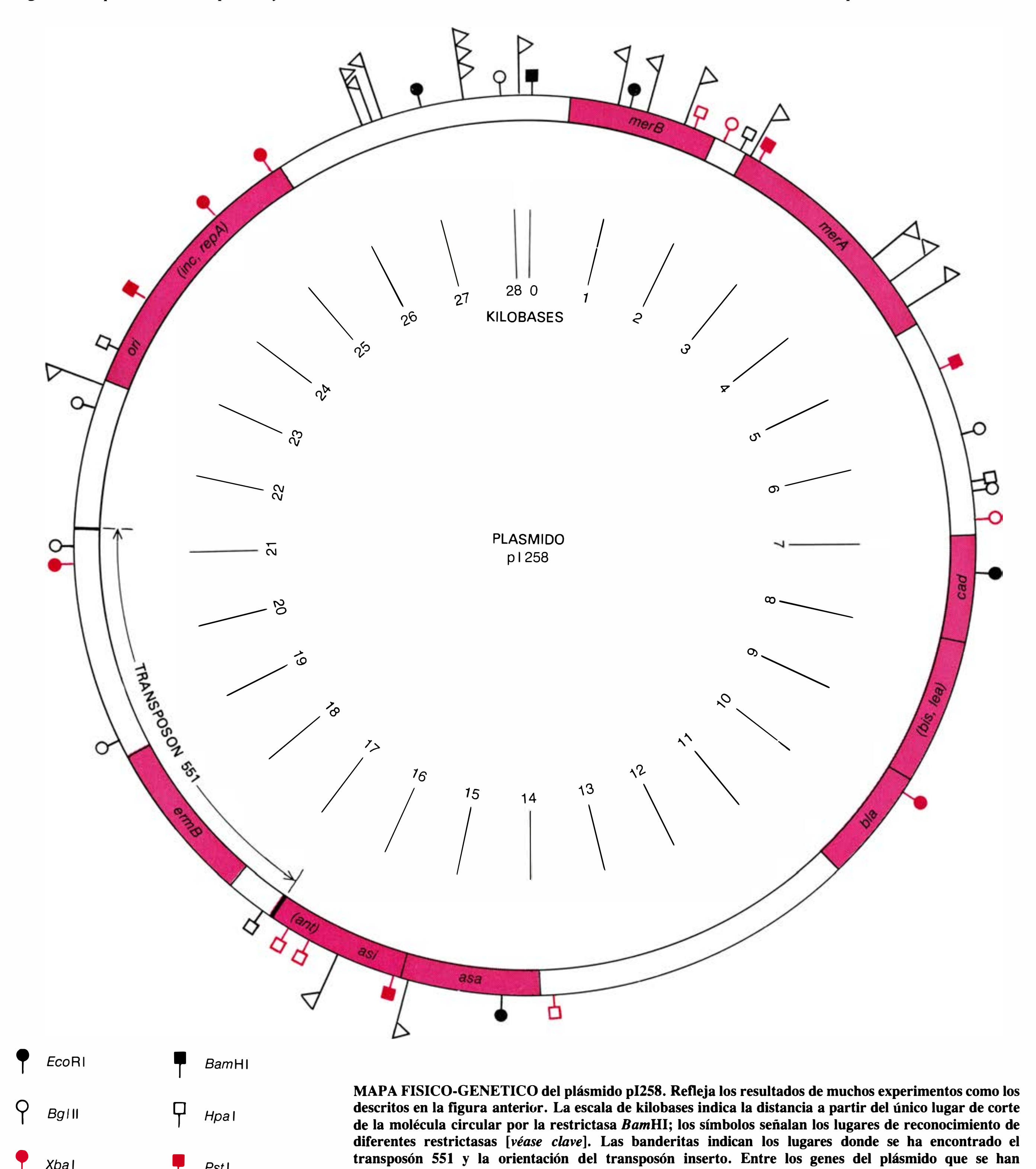
Bloque replicador

Recientes experimentos de clonación molecular han definido mejor los bloques replicadores de varios plásmidos de *E. coli*. En estos experimentos se digiere un plásmido con una restrictasa y se mezcla el conjunto de fragmentos resultantes con un fragmento "seleccionador" de ADN, generado por la misma restrictasa, incapaz de replicarse si bien dotado de un marcador seleccionable. Puesto que todos los fragmentos han sido cortados por la misma restrictasa, tienen en sus extremos bases complementarias simétricas y no apareadas ("extremos cohesivos"); por apareamiento entre estas bases pueden combinarse unos con otros en cualquier número, orden y orientación relativa. Cuando se trata la mezcla con el enzima ADN ligasa, los fragmentos enlazados forman nuevos plásmidos según reordenaciones aleatorias del plásmido original. Al introducir estos plásmidos en células receptoras, sólo se replicarán aquellos que contengan el segmento que les confiere autonomía, y únicamente podrán recuperarse los que porten el fragmento seleccionador. Se pueden determinar los fragmentos que componen el nuevo plásmido tratándolo con la restrictasa empleada en primer lugar; su orden y orientación relativa se deducirán utilizando una restrictasa distinta.

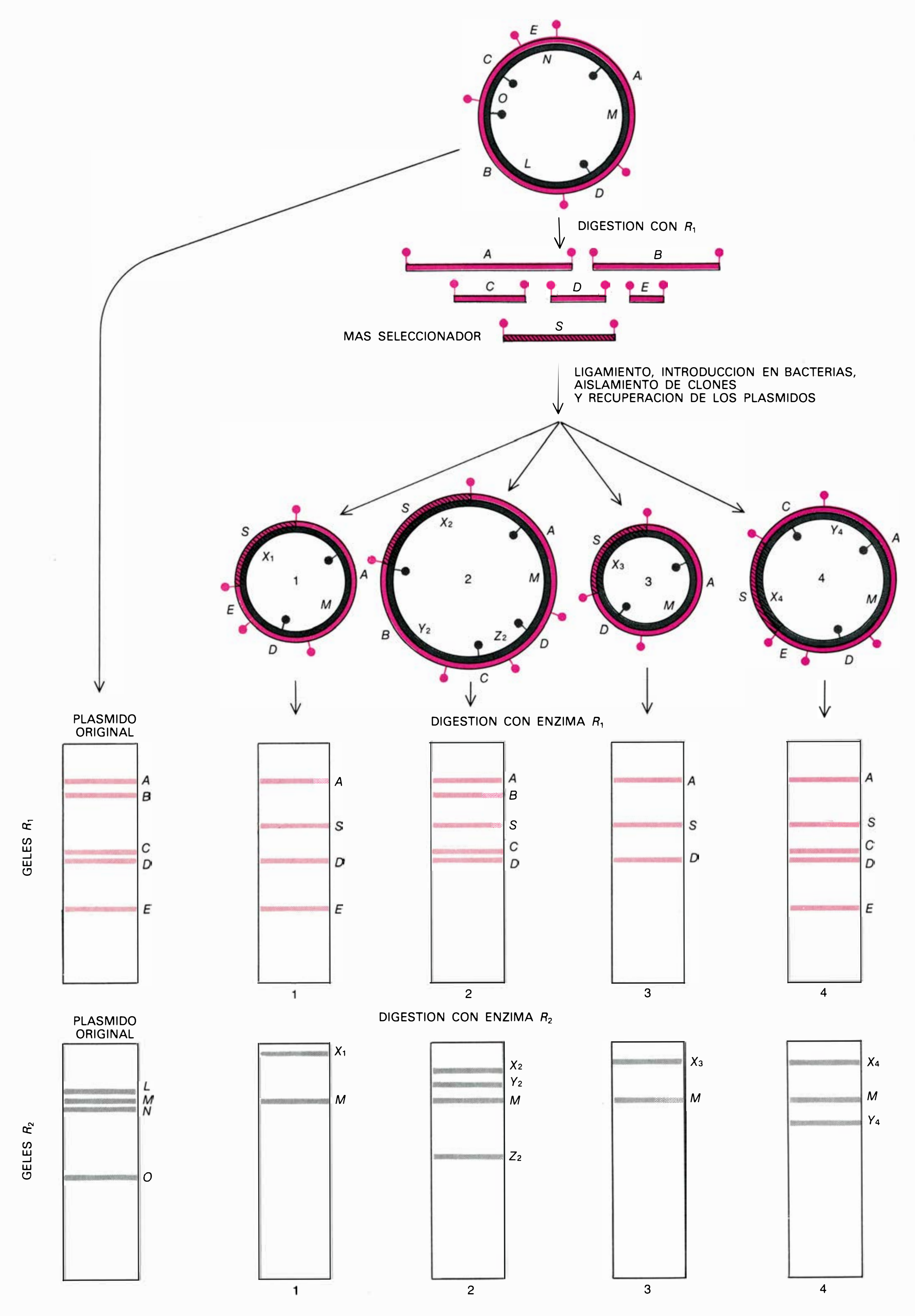
Si uno de los fragmentos originales contenía todos los genes responsables de la replicación, ese fragmento se encontrará en todos los nuevos plásmidos viables. Kenneth N. Timmis, Felipe Cabello y Stanley N. Cohen, de la Facultad de Medicina de la Universidad de Stanford, comprobaron que éste era el caso en dos grandes plásmidos de *E. coli*, el F y el R6. En ambos, un solo fragmento, unas diez veces menor que el plásmido original, contenía todo el bloque replicador y permitía una replicación autónoma equiparable a la del plásmido original entero. Nuevas reducciones de tamaño de los plásmidos mutantes indicaron que un segmento de ADN de no más de 2000 pares de bases albergaba todo el bloque, capaz de replicar plásmidos más de 50 veces mayores que él.

Si eran dos los fragmentos que contenían las funciones necesarias para la replicación, ambos se encontraban siempre en los nuevos plásmidos, aunque no en situación relativa constante. Si un punto de disección de la restrictasa se encontraba en un gen necesario, se recobraban dos de los fragmentos originales, que eran siempre adyacentes y guardaban la misma orientación que en el plásmido original.

Se han definido experimentalmente, hasta ahora, cuatro elementos de un bloque replicador típico: el origen (lugar de iniciación de la replicación), el iniciador, las funciones genéticas que determinan el número de copias y las que aseguran el reparto de las copias a las células hijas. El elemento central del sistema autonómico es el origen de replicación, una secuencia específica de bases en la que se encuentra el sitio real de iniciación de cada ciclo de replicación. El origen se define funcionalmente como la menor secuencia de ADN que puede servir de punto de partida de un ciclo de replicación.



MAPA FISICO-GENETICO del plásmido pI258. Refleja los resultados de muchos experimentos como los descritos en la figura anterior. La escala de kilobases indica la distancia a partir del único lugar de corte de la molécula circular por la restrictasa *Bam*HI; los símbolos señalan los lugares de reconocimiento de diferentes restrictasas [*véase clave*]. Las banderitas indican los lugares donde se ha encontrado el transposón 551 y la orientación del transposón inserto. Entre los genes del plásmido que se han cartografiado se encuentran los responsables de la resistencia a la penicilina (*bla*), arseniato (*asa*), arsenito (*asi*), antimonio (*ant*), eritromicina (*ermB*), mercurio inorgánico (*merA*), mercurio orgánico (*merB*), cadmio (*cad*), bismuto (*bis*) y plomo (*lea*), el origen de replicación (*ori*), el iniciador (*repA*), y la incompatibilidad (*inc*). Aquellos genes cuya localización no es segura se muestran entre paréntesis.



DIGESTION CON R_1
MAS SELECCIONADOR
LIGAMIENTO, INTRODUCCION EN BACTERIAS, AISLAMIENTO DE CLONES Y RECUPERACION DE LOS PLASMIDOS
PLASMIDO ORIGINAL
DIGESTION CON ENZIMA R_1
GELES R_1
PLASMIDO ORIGINAL
DIGESTION CON ENZIMA R_2
GELES R_2

Una variante del experimento de reordenación de fragmentos permite aislar el origen de replicación. Como antes, se une una mezcla de fragmentos producidos por restrictasas con un fragmento seleccionador. Los nuevos plásmidos se introducen ahora en una cepa bacteriana, diseñada especialmente para el caso, que contiene un segundo plásmido no relacionado, en el que se ha insertado un fragmento del primero que codifica para todas las sustancias difusibles necesarias para su replicación, como el iniciador. También aquí, el experimento señala el fragmento indispensable para que el plásmido reordenado sea viable. En este caso, sin embargo, el fragmento requerido no tiene por qué llevar todo el bloque replicador, pues las sustancias difusibles necesarias podrá suministrarlas el plásmido complementario ya presente en las células. Por consiguiente, el fragmento que esté siempre presente en los nuevos plásmidos viables deberá contener específicamente el ADN del origen de replicación. De este modo, Donald R. Helinski y sus colegas aislaron los orígenes de los plásmidos R6K y RK2 de *E. coli,* y con métodos parecidos se han aislado los orígenes de otros plásmidos.

En todos los casos, el origen tenía varios cientos de pares de bases. Dentro de esta región, la síntesis de la nueva cadena de ADN comenzaba prácticamente siempre en el mismo par de bases. El resto del origen de replicación contiene posiblemente secuencias específicas a las que se unen diversas proteínas, y quizá otras macromoléculas, para influir sobre el proceso de iniciación. Por ejemplo, parece que el primer paso de la replicación del ADN plasmídico (y de otros) es el recorrido de la secuencia origen por la ARN polimerasa. Esta interacción no sólo sintetiza una cadena de ARN que permite "cebar" la replicación del ADN, sino que separa las dos cadenas de ADN, con lo que quizá se produzcan apareamientos intracatenarios parciales que podrían ser reconocidos por los enzimas y otras proteínas implicadas en la biosíntesis del nuevo ADN.

La primera señal de que los plásmidos contenían genes necesarios para su replicación salió del trabajo de Jacob, Brenner y Cuzin. Aislaron mutantes del plásmido F inestables a temperaturas elevadas, al parecer por no poder replicarse. Nosotros aislamos mutantes similares de varios plásmidos de estafilococos y demostramos que, en efecto, no se replicaban a temperaturas elevadas, ambiente en el que sus huéspedes se desarrollaban sin problemas. Estos plásmidos eran capaces de replicarse cuando la célula contenía copias del mismo plásmido que llevaban la versión (alelo) normal activa del gen que había sufrido la mutación responsable de la termosensibilidad. Ello significaba que este gen determinaba una proteína difusible (que se supone el iniciador) y que el plásmido normal suministraba esta proteína al plásmido mutante. Los mutantes no consiguieron replicarse en presencia de un plásmido diferente, aunque estrechamente emparentado; se reflejaba así una alta especificidad en la interacción entre la proteína y el ADN y se cumplía la predicción de que existía un iniciador específico para cada replicación. Manabu Inuzuka y Helinski han demostrado que el iniciador del plásmido R6K es necesario para la replicación del plásmido en extractos crudos, es decir, libres de células, de *E. coli*.

EXPERIMENTO DE REORDENACION de fragmentos. Dicha prueba permite definir la región responsable de las funciones autonómicas del plásmido. En el diagrama del plásmido (*arriba*) se indican los lugares de corte de dos hipotéticas restrictasas, R_1 (*señales coloreadas*) y R_2 (*grises*). El plásmido se digiere primero con R_1 y los fragmentos ($A - E$) se mezclan con el fragmento seleccionador (*S*), dotado de extremos cohesivos; se unen entre sí, y los nuevos plásmidos resultantes se clonan en bacterias. Los plásmidos viables (los que contienen las funciones autonómicas) se replican y se pueden seleccionar si llevan además el fragmento seleccionador. La digestión de los plásmidos reordenados con R_1 revela un conjunto diferente de fragmentos en cada plásmido, pero todos tienen los fragmentos *S, A* y *D;* los dos últimos deben poseer, entre ambos, todas las funciones autonómicas. La disección del plásmido original con el enzima R_2 da un conjunto diferente de fragmentos, *L, M, N* y *O;* los plásmidos reordenados contienen todos *M* y varios fragmentos nuevos. Como *M* se halla en todos los plásmidos viables, tiene que contener funciones autonómicas; la constancia de su presencia muestra que *A* y *D* son adyacentes y guardan la misma orientación relativa en todas las reordenaciones viables.

¿Cuánta replicación?

Se supuso al principio que la célula tendría tantas copias de cada plásmido cuantos cromosomas y que la replicación del plásmido estaría acoplada con, y regulada por, la del cromosoma. Pero no es así: en los plásmidos estudiados hasta ahora, el número de copias por célula parece variar sin solución de continuidad desde uno o dos hasta más de 100; en general, hay menos copias cuanto mayor es el plásmido. Y la replicación del plásmido suele estar regulada independientemente de la del cromosoma. Dado que los distintos plásmidos que se encuentren en una célula pueden estar representados por un número de copias muy diferente, este número debe estar determinado por un sistema regulador propio del plásmido. El primer sistema de este tipo fue descubierto en el plásmido R1 de *E. coli,* en 1972, por Kurt Nordström, que ahora está en la Universidad de Odense, en Dinamarca, y sistemas parecidos se han descubierto en otros plásmidos de estafilococos y de *E. coli.* En R1, por lo menos, el número de copias parece estar controlado por una proteína o proteínas que inhiben la replicación, con lo que se cumple una predicción hecha en 1968 por Robert H. Prichard, Peter T. Barth y John F. Collins, de Leicester, de que la replicación está controlada negativamente, es decir, por un represor.

¿Cuándo se replican los plásmidos en el ciclo celular y cómo se determina el orden en que se replican las copias de un plásmido determinado? Pudiera haber un momento preciso en que todas las moléculas de plásmidos se replicaran sincrónicamente; las pruebas sugieren, en cambio, que los plásmidos se replican aleatoriamente durante todo el ciclo celular. Un elegante experimento publicado en 1967 por Robert H. Rownd, de Wisconsin, demostró que las copias de cada plásmido formaban un grupo homogéneo, en cuanto que se elegían aleatoriamente para su replicación, sin tener en cuenta sus replicaciones recientes; el sistema regulador no distinguía entre moléculas "nuevas" y "viejas" de ADN. Otros investigadores han puesto de manifiesto que los plásmidos se replican durante todo el ciclo celular. Una sencilla interpretación de estas observaciones sería que, a cada velocidad de crecimiento del huésped, corresponde una concentración constante de represor que determina la probabilidad de replicación del plásmido.

No está todavía claro cómo interaccionan los dos elementos, el iniciador, positivo, y el represor, negativo, para regular la replicación. Posiblemente difiera entre sistemas distintos. Un represor podría regular la velocidad de síntesis del iniciador, la afinidad de éste con el origen, o la frecuencia de transcripción del origen por la ARN polimerasa. Cualquiera de estos mecanismos se aparta del predicho por Jacob, Brenner y Cuzin, que implica formalmente la presencia de sólo un regulador positivo y de un número limitado de lugares a los que tiene que unirse el replicón para multiplicarse.

Jacob, Brenner y Cuzin advirtieron que la estabilidad hereditaria de un elemento genético no esencial representado en cada célula por un número limitado de copias exigía la existencia de un mecanismo para el reparto iguali-

tario de copias a las células hijas. Tal mecanismo no se ha identificado todavía con claridad, pero la primera demostración directa de su existencia procede de un estudio de nuestro laboratorio sobre plásmidos mutantes que tienen un iniciador termosensible. Al dividirse el huésped a temperatura elevada, los plásmidos presentes no se destruyen ni se replican; se reparten entre las células hijas hasta que ninguna célula tiene más de un plásmido.

Reparto

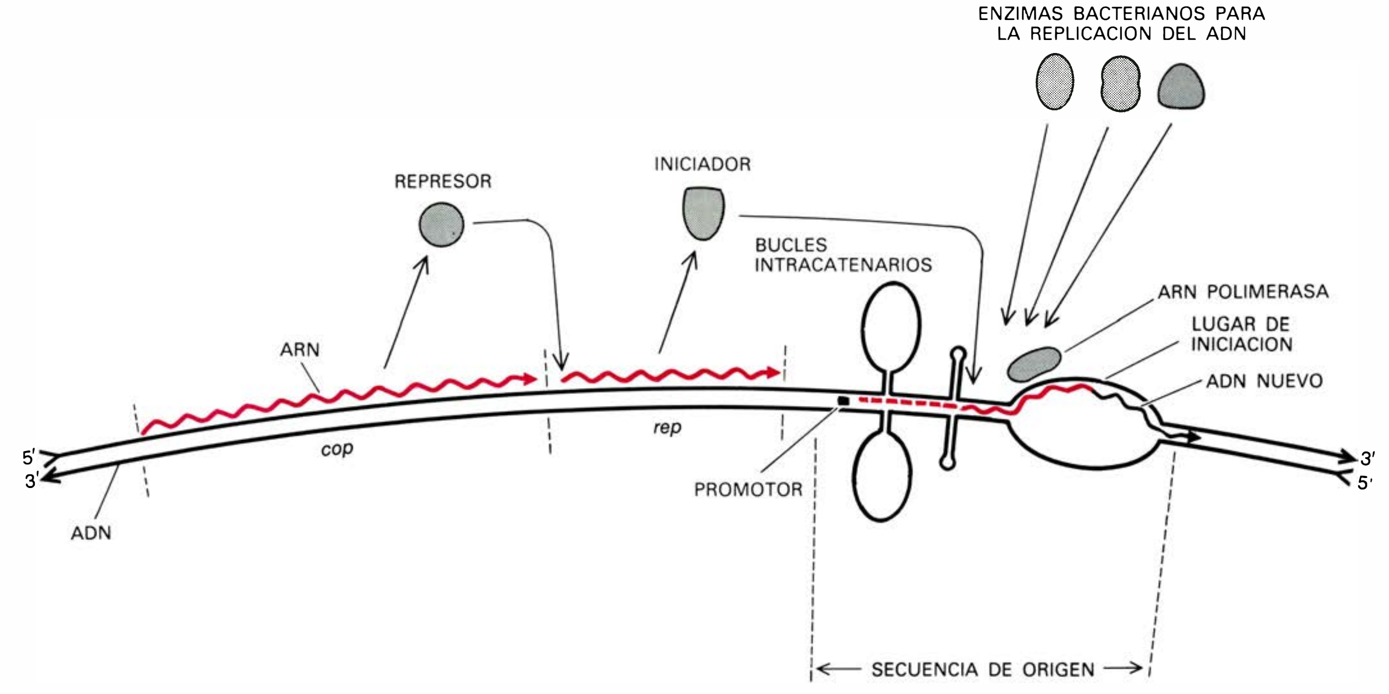
Estudiamos estadísticamente este reparto en el caso de dos plásmidos diferentes, uno con una media de 2 a 3 copias por célula y otro con unas 30 copias por célula. En ambos casos, el número de copias se redujo aproximadamente a la mitad en cada generación hasta que sólo quedó una por célula. En cada división posterior, una célula hija recibía un plásmido y la otra ninguno. Concluimos que por lo menos estos dos plásmidos tenían un mecanismo específico de reparto igualitario. Probablemente el mecanismo implicaba a la membrana celular y constaba también de dos componentes, una molécula efectora y un lugar de reconocimiento.

Nordström y Rownd han observado independientemente hace poco que, cuando se elimina cierto fragmento de un plásmido R por técnicas de ADN recombinante, el plásmido se vuelve inestable y se hereda de modo aparentemente aleatorio; la interpretación provisional es que el segmento eliminado tiene que ver con el reparto equitativo. En colaboración con Carmen Sánchez Rivas, de la Universidad de París, hemos estudiado plásmidos en protoplastos de estafilococos, es decir, bacterias a las que se han quitado sus fuertes paredes celulares, dejando sólo la fina membrana celular. Al regenerar los protoplastos una pared celular, muchas de las células regeneradas no heredan los plásmidos que se hallaban en un origen. Ahora bien, el ADN de los plásmidos no se expulsa de los protoplastos y se puede replicar normalmente; por tanto, el fallo debe estar relacionado con las divisiones celulares que ocurren durante la regeneración de la pared celular. Al ser la pared celular la que determina la situación del plano de división, estas divisiones son anormales y dan lugar a menudo a "células" sin cromosoma (que por supuesto no pueden sobrevivir) y a células sin plásmidos.

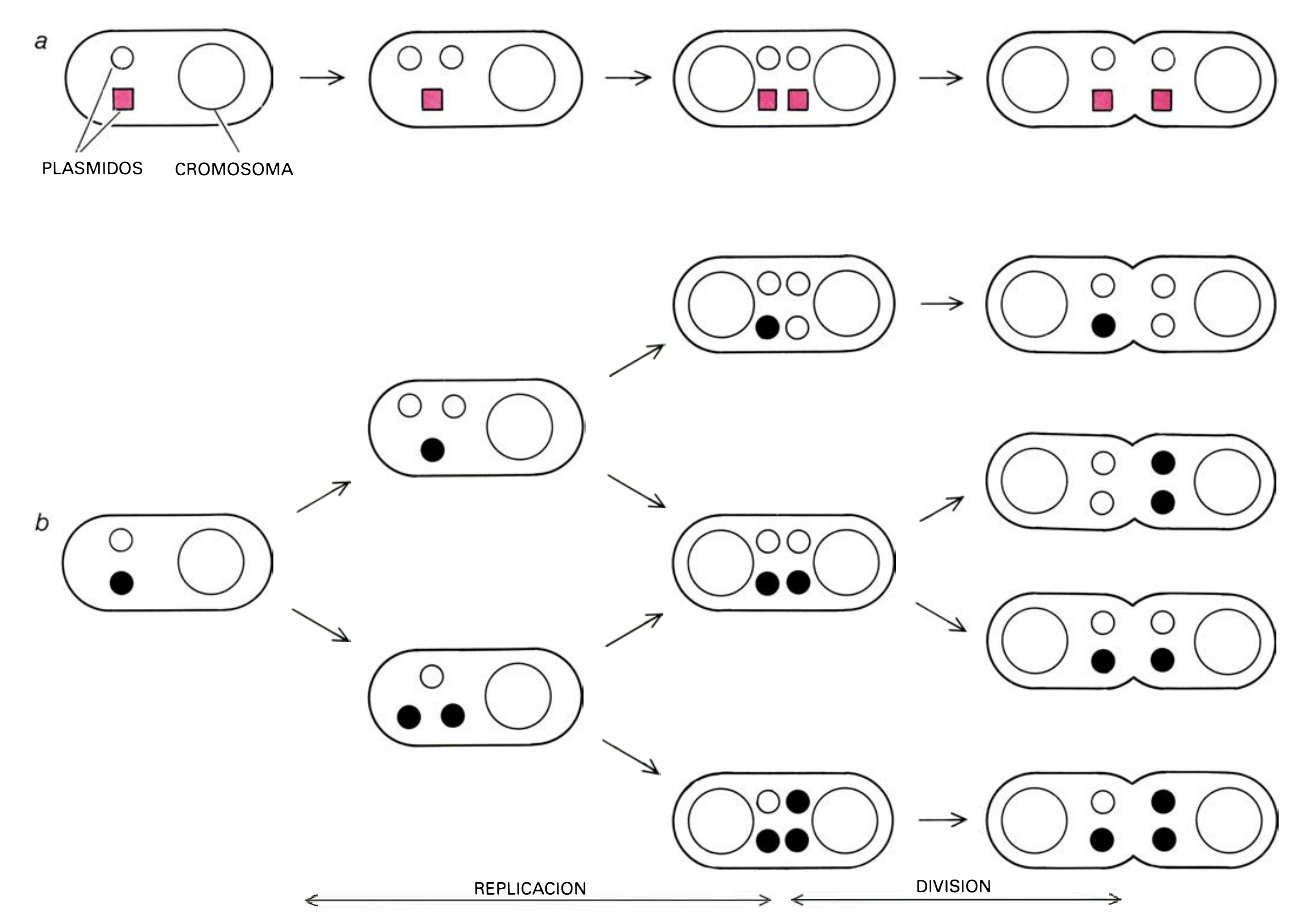
Este resultado indica que la pared celular no alterada está íntimamente relacionada con la coordinación de la división y el reparto, de acuerdo con la teoría de Jacob, Brenner y Cuzin. La teoría especifica, sin embargo, que, tras la replicación, las dos copias de un replicón se separan automáticamente, yendo una a cada célula hija. Sabemos ya que esto no puede ser verdad en el caso de los plásmidos a causa de un fenómeno exclusivo de ellos: la incompatibilidad.

Incompatibilidad

Los plásmidos de parentesco afín no suelen coexistir establemente en una misma célula: son "incompatibles". Tal vez se puedan obtener células que contengan los dos plásmidos, pero es difícil mantenerlos juntos. Al proliferar las células, se generan continuamente subclones que sólo tienen uno u otro de ambos plásmidos; el cultivo llega con el tiempo a contener dos sublíneas puras, cada una con uno de los plásmidos originales. Frank C. Hoppensteadt, de la Universidad de Utah, y yo hemos demostrado, mediante análisis estadístico de experimentos realizados en nuestro laboratorio y en el de Nordström, que el proceso era sin duda estocástico: los resultados pueden interpretarse como el efecto neto de una replicación y reparto aleatorios. Dos plásmidos incompatibles A y B presentes en una misma célula forman un grupo homogéneo del que se eligen aleatoriamente las copias que se van a replicar, hasta que se ha duplicado el número total; se eligen luego, también de forma aleatoria, las dos mitades que van a tocar a las células hijas en el reparto. Las proporciones de A y B variarán de unas células a otras según una distribución predecible matemáticamente, aunque la suma de A y B permanecerá constante. Puesto que los



LA REPLICACION DE LOS PLASMIDOS puede estar regulada según este esquema. La actividad del gen *rep* (por replicación) está regulada por una proteína represora que viene especificada por el gen *cop* (por número de copias). La proteína iniciadora reconoce bucles en las dos cadenas del ADN del plásmido. Estos bucles o lazos se forman por apareamiento de bases de una misma cadena al separarse las cadenas mientras la ARN polimerasa recorre la secuencia origen a partir del promotor para sintetizar un ARN que servirá de cebo para la síntesis de ADN. El complejo formado por los bucles y el iniciador atrae a diversos enzimas de la célula huésped, entre ellos la ADN polimerasa, la cual añade nucleótidos (*negro*) al cebo de ARN (*color*).



REPLICACION Y REPARTO de plásmidos compatibles (*a*). Ambos procesos contrastan con los de los plásmidos incompatibles (*b*). Los plásmidos no relacionados son "compatibles", es decir, pueden coexistir en la misma célula. Ambos se replican, posiblemente en distintos momentos del ciclo celular, y sus copias se reparten por igual durante la división celular. Los plásmidos muy emparentados son "incompatibles". Sus copias se eligen al azar para ser replicadas (*columnas centrales*) y se reparten al azar (*derecha*). El resultado inevitable es la producción de células que carecen de uno u otro plásmido.

dos plásmidos tienen bloques replicadores parecidos o idénticos, los mecanismos que determinan la replicación y el reparto no distinguen un plásmido del otro. Persisten, por tanto, las desproporciones y aparecen, a una velocidad predecible a partir del número total de copias, células que tienen sólo *A* o sólo *B*.

A diferencia de los plásmidos dotados de bloques replicadores parecidos o idénticos pueden coexistir en una célula pares de plásmidos que no guardan relación entre sí; pueden considerarse incluso compatibles. Las bacterias salvajes tienen frecuentemente hasta siete u ocho plásmidos diferentes. El contraste entre compatibilidad e incompatibilidad indica un alto grado de especificidad en las interacciones entre células y plásmidos. Los plásmidos poseen sin duda genes específicos responsables de la diferencia entre plásmidos compatibles e incompatibles. Quizás el reparto sea específico para cada tipo de incompatibilidad, es decir, los plásmidos incompatibles utilizan el mismo sistema de reparto. En tal caso, habría por lo menos tantos sistemas de reparto cuantos grupos de plásmidos mutuamente incompatibles; se ha identificado ya un número considerable de tales grupos de incompatibilidad en varias especies bacterianas. Todo ello sugiere que la especificidad del reparto está determinada exclusivamente por el plásmido.

El estudio del reparto y la incompatibilidad de los plásmidos nos llevan de la mano hasta la esencia de la simbiosis. Los endosimbiontes subcelulares utilizan, a buen seguro, las estructuras, vías metabólicas y otras funciones del huésped con la misma finalidad que lo hace el propio huésped. Así, un plásmido utiliza presumiblemente la ARN polimerasa de su huésped para transcribir su propio ADN y las ADN polimerasas de su huésped para replicar y reparar su propio ADN. De forma alternativa, los organismos subcelulares pueden subvertir las funciones de su huésped para atender necesidades peculiares propias. Por ejemplo, el sistema por el que la célula bacteriana transporta el azúcar maltosa a través de su membrana es utilizado por el fago lambda para una finalidad muy diferente: como lugar de unión a la pared de la célula que infecta. En otros casos, tales como la regulación del número de copias, puede que no haya una función celular apropiada que pueda utilizarse y el plásmido mismo tiene que determinar las moléculas reguladoras necesarias.

¿Qué decir del reparto? A efectos del cual, muchos plásmidos distintos tienen que haber desarrollado independientemente una interacción estructural específica con su huésped. Supongo que cada plásmido quizá haya "aprendido" a anclarse en una estructura celular diferente (tal vez en el interior de la membrana celular o incluso en el propio cromosoma, como ha sugerido Bruce C. Kline, de la Clínica Mayo), desarrollando por evolución una proteína que se una a un lugar específico. De ocurrir así, el aislamiento y el estudio de tales sistemas ayudaría a clarificar ese tipo especial de simbiosis molecular que caracteriza la forma de vida de los plásmidos.

El hombre primitivo en el Asia central soviética

En la República Socialista Soviética de los Tadjiks fueron acumulándose depósitos de loess a lo largo de dos millones de años. Incluidos en ellos se hallan instrumentos de piedra tallados por grupos cazadores del Paleolítico inferior

Richard S. Davis, Vadim A. Ranov y Andrey E. Dodonov

No hay pruebas de la presencia del hombre primitivo en gran parte de Asia. Está claro que fue allí donde se descubrió por vez primera un *Homo erectus* (el "hombre de Java") y también en Asia donde se excavó el yacimiento de Chukutien, que reveló una enorme cantidad de restos esqueléticos de la misma especie (el "hombre de Pekín"). Sin embargo, extensas zonas de ese continente constituyen espacios vacíos en los mapas arqueológicos. Por ejemplo, hasta hace pocos años la región extendida al este del Eufrates por toda el Asia central, a través de la meseta del Irán, carecía prácticamente de yacimientos del hombre del Paleolítico inferior que pudiesen ser fechados con seguridad en relación con su contexto geológico. Por ello era poco lo que se conocía de los cazadores primitivos y sus modos de vida en esa ancha zona árida del Asia interior.

Recientemente, en la república soviética del Tadjikistán, del Asia central, se descubrieron diversos yacimientos arqueológicos, recubiertos por más de 60 metros de espesor de loess transportado por el viento. Por indicios diversos, esas huellas de actividad humana pueden situarse hace aproximadamente un cuarto de millón de años, época que corresponde al intervalo más reciente del Paleolítico inferior. Pero, además de su contenido arqueológico, los depósitos de loess de Tadjikistán poseen otro tipo de interés. Permiten probar, por ejemplo, la realidad de una secuencia aparentemente continua de oscilaciones climáticas, cálida a fría, extendida a lo largo de los dos últimos millones de años, es decir, el Pleistoceno entero. La conjugación de los dos tipos de información que posee este loess, la arqueológica y la climática, nos proporciona una visión de las adaptaciones del hombre primitivo en un medio continental de tierras altas durante los tiempos del Paleolítico inferior.

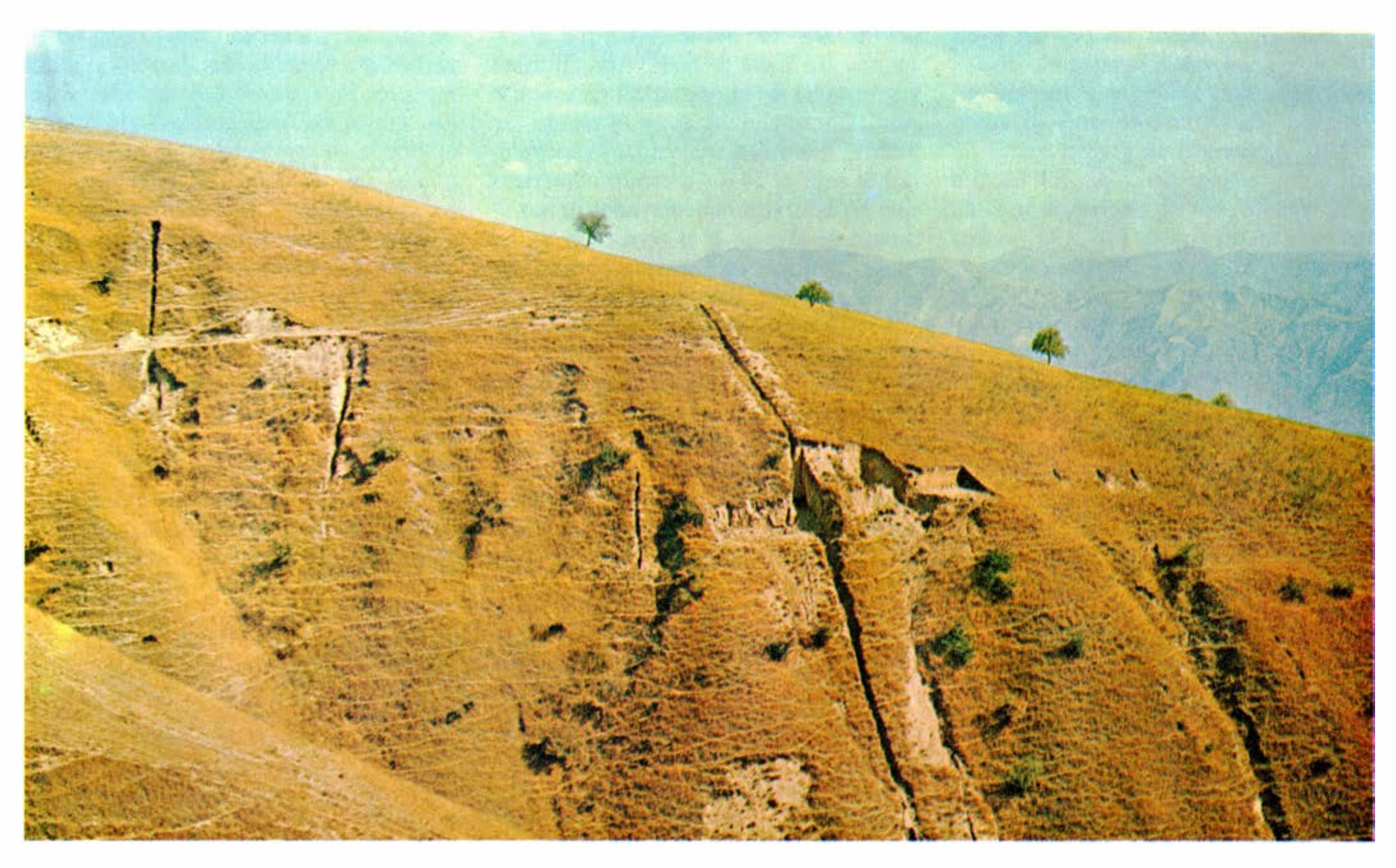
El loess (del alemán helvético lösch, "suelto, poroso") es en principio un sedimento transportado por el viento. Típicamente amarillento, calcáreo y poroso, está compuesto por partículas de tamaño limo. En las regiones exentas de hielos se forma ordinariamente bajo condiciones bastante secas, en áreas cuyo suelo no está suficientemente protegido por la vegetación, por lo que su superficie se erosiona con facilidad.

Depósitos espesos de loess existen a lo largo del río Amarillo en China, en la llanura de Potwar en Pakistán, en Ucrania, en Europa central y en la cuenca del alto Mississippi. Los depósitos a lo largo de la frontera entre Afganistán y Tadjikistán no son tan bien conocidos, pero constituyen probablemente una de las mayores áreas de loess del mundo. Aunque muchos de estos depósitos no están consolidados, no por ello se desploman fácilmente; las aguas de arrollada, por ejemplo, producen en ellos escarpes casi verticales. Al sur de Tadjikistán son frecuentes estas entalladuras profundas, y favorecen el estudio de los cambios del Pleistoceno a lo largo del registro sedimentario.

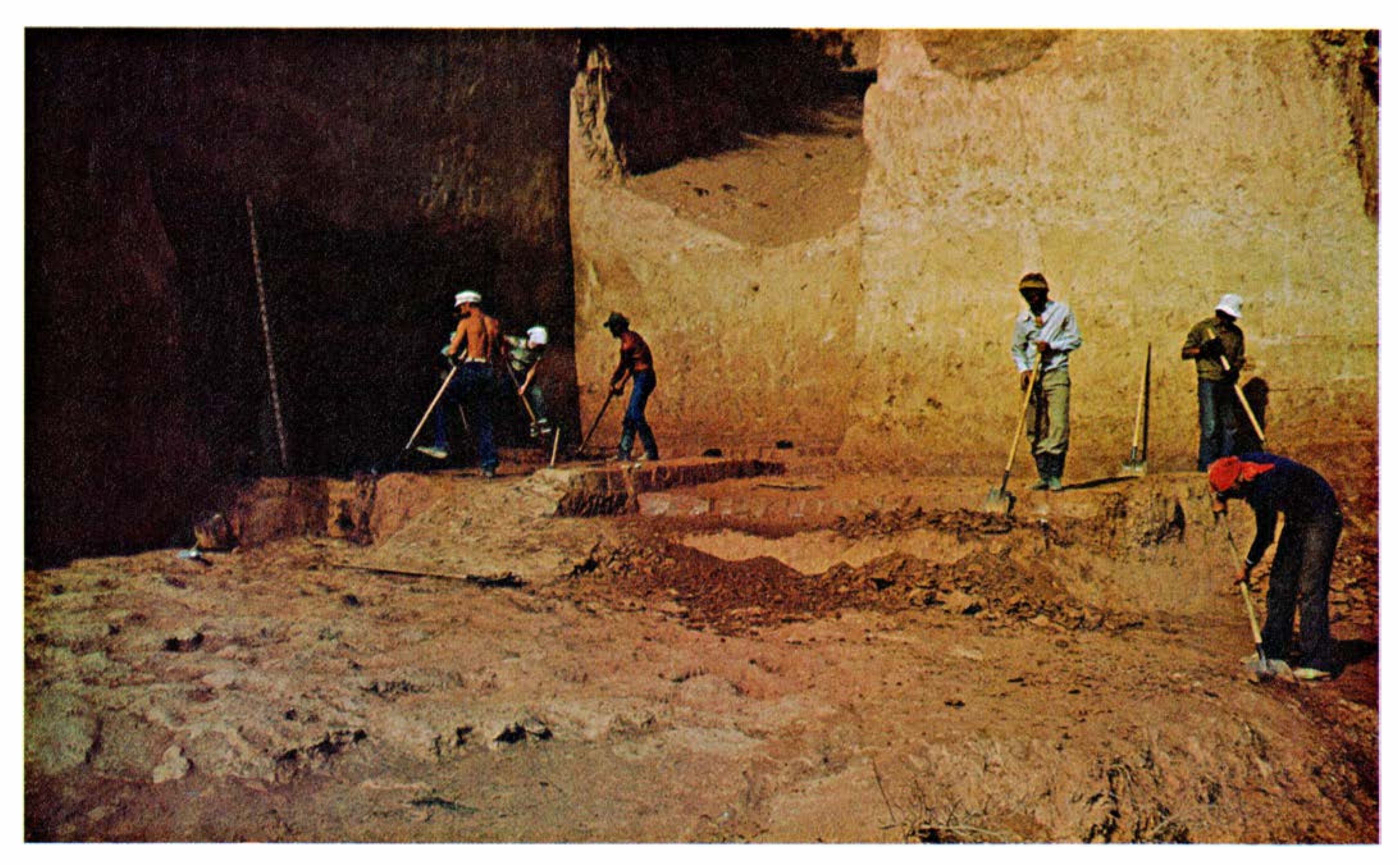
Dentro del Asia central soviética, el Tadjikistán abarca un área montañosa tectónicamente activa originada por el levantamiento y fracturación de las alineaciones de Gissar, Pamir e Hindokush durante el Terciario superior, como consecuencia de la colisión por deriva del subcontinente de la India con Asia. Una zona de subsidencia relacionada con estos levantamientos es la depresión afgano-tadjik. Durante millones de años, la depresión actuó de área de depósito de los materiales arrancados de las cadenas montañosas que la rodean. En diversas partes de la depresión se formaron, durante el Pleistoceno, espesos depósitos de loess. En algunos puntos poseen 100 metros de potencia y en otros superan los 200.

Hacia 1960 se iniciaron investigaciones, a gran escala, de este loess por parte de estudiosos del Instituto de Geología de la Academia de Ciencias de la Unión Soviética y de la del Tadjikistán. El trabajo no ha finalizado todavía y los materiales loéssicos se están examinando por especialistas en clima, en suelo y formaciones sedimentarias, en minerales y en polen. La edad de los depósitos puede establecerse por datación paleomagnética de las finas partículas de magnetita, el mineral ferruginoso presente en el loess, por datación mediante termoluminiscencia de partículas de cuarzo (uno de los componentes más importantes del loess) y por correlación estratigráfica. Arqueólogos y paleontólogos estudian además el utillaje y los restos fósiles que han sido descubiertos. Todo ello está proporcionando ya resultados interesantes.

Un rasgo que inmediatamente llama la atención en los afloramientos del loess del Tadjikistán lo constituye la alternancia de bancos espesos de loess sin alterar con varios "horizontes" de suelo. Los horizontes edáficos, o de suelo, se formaron cuando la superficie del loess sufría alteración durante períodos de clima relativamente húmedo y cálido. En varios puntos, la estructura del suelo es compleja y se superponen dos o hasta tres horizontes; ello indica la interacción dinámica de dos procesos simultáneos: sedimentación y formación del suelo. Típicamente los



LADERAS ESCARPADAS EN EL LOESS de Karatau en Tadjikistán, que descienden más de 1100 metros hasta el río Vaksh. Cuatro estrechos tajos son visibles a la izquierda de un surco más amplio que penetra en la excavación (*a la derecha, junto al centro*), mostrando un complejo de suelos donde se extrajeron los instrumentos líticos. El suelo, que se encuentra actualmente a 67 metros bajo la superficie, debió desarrollarse hace unos 250.000 años.



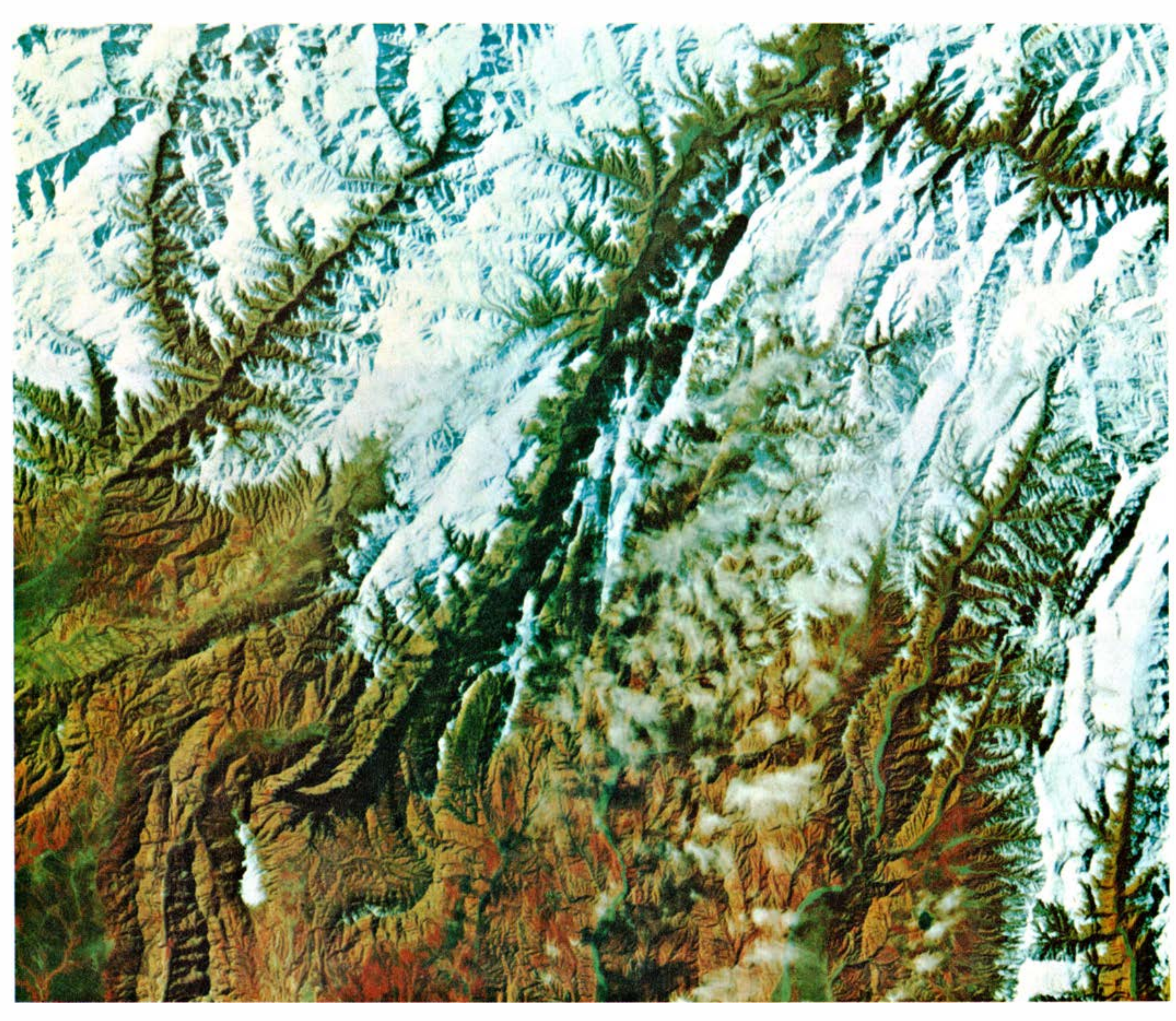
LA EXCAVACION EN LAKHUTI descubrió otro suelo fósil que contenía instrumentos líticos. Era el quinto de los nueve complejos edáficos en Lakhuti I, todos ellos situados por encima del nivel donde, en los depósitos de loess del Tadjikistán, los datos paleomagnéticos indican el cambio de la polaridad magnética de hace 700.000 años (Matuyama) a la polaridad normal de la actualidad (Brunhes). Aquí se encontraron hasta 450 instrumentos líticos.

paleosuelos o suelos fósiles del loess son pardo-rojizos, contrastando con el color amarillento de esta formación; alcanzan espesores de tres a cinco metros y poseen una gruesa costra de carbonato cálcico cerca de la base. La zona más intensa de alteración se sitúa ordinariamente hacia los 1,50 metros de suelo por encima de la costra.

Las propiedades físicas y químicas de los paleosuelos del loess son análogas a las de los suelos castaños (*kastanozemy,* en terminología rusa) que en la actualidad se originan bajo condiciones de estepa semiárida. El polen vegetal y las especies de caracoles encontradas en el loess indican, sin embargo, que éste se acumulaba bajo un clima considerablemente más frío y seco que el actual. De este modo, las capas alternantes de loess y suelo constituirían una prueba de la existencia de amplias oscilaciones climáticas en la región. Se admite que tales oscilaciones representan el reflejo de los cambios climáticos que en latitudes más elevadas determinaban el avance y el retroceso de las masas de hielo en las áreas continentales durante el Pleistoceno.

Las dataciones paleomagnéticas están basadas en el hecho de que los polos N y S magnéticos de la tierra se invierten durante períodos de decenas o centenas de millares de años. Como las finas partículas de magnetita se depositaron junto con el loess, tienden a orientarse de acuerdo con la polaridad existente en la época, y esta orientación puede medirse con un magnetómetro sensible. La datación de las mayores inversiones de polaridad y de los episodios más breves de dicha inversión se establecen independientemente de la datación radiactiva de rocas volcánicas. Por consiguiente, es posible, una vez reconocida la secuencia de las inversiones en el loess, señalar una edad correcta para los estratos identificados.

En el loess de Tadjikistán pueden detectarse diversos episodios paleomagnéticos, el más importante de los cuales lo constituye el paso, hace unos 690.000 años, del período Matuyama, cuando la polaridad magnética terrestre tenía sentido opuesto a la de hoy, al actual período Brunhes. En las seccio-



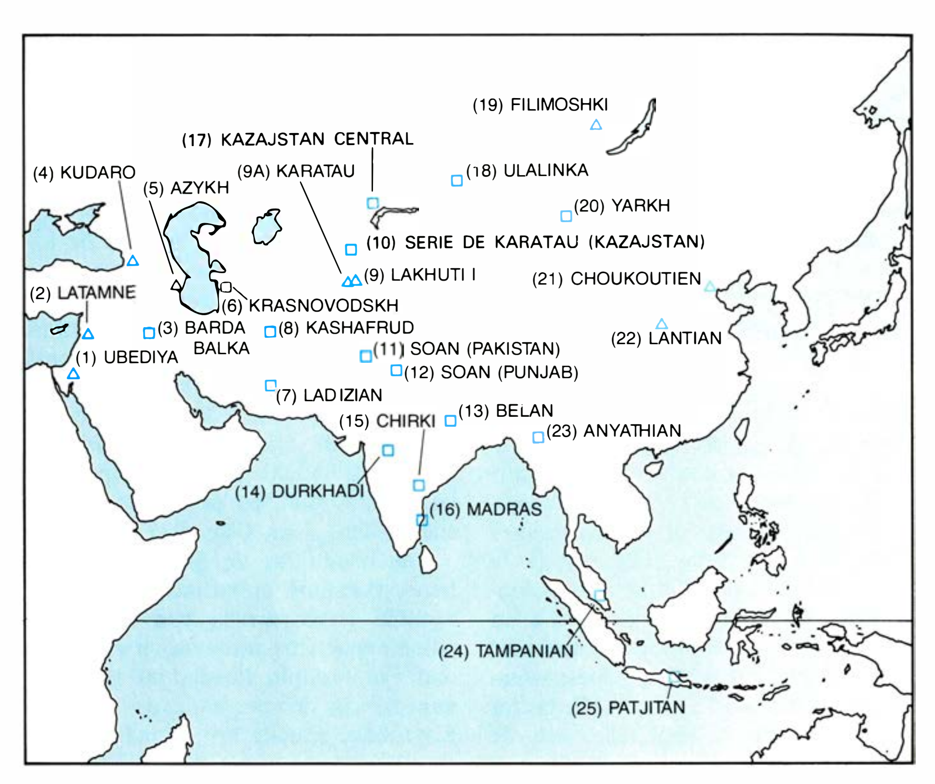
EXTENSOS DEPOSITOS DE LOESS subyacentes al límite de las nieves de mediados de febrero, en el área fronteriza entre Afganistán y Tadjikistán, al este de Dushanbe. La imagen vía satélite fue obtenida por el Centro de Datos EROS del Servicio Geológico de Estados Unidos. Los datos para la imagen se tomaron desde una altura de 915 kilómetros por el satélite para recursos terráqueos *Landsat 2* durante el invierno 1978-79. Los depósitos de loess profundamente entallados en la zona fueron estudiados en las postrimerías de la década de 1960. Se han recogido pruebas de la presencia del hombre primitivo en el rudo ambiente de esta región esteparia en 12 secciones excavadas por la erosión; los datos recibieron confirmación arqueológica de los trabajos efectuados en dos de dicha secciones. El área cubierta por la imagen aparece encuadrada en el mapa inferior de la página siguiente.

nes de loess del Tadjikistán que actualmente son objeto de estudio, el número total de complejos edáficos existente varía, siendo de 37 el máximo. En los seis perfiles en que el límite Matuyama-Brunhes ha podido ser señalado, nueve de estos complejos edáficos figuran por encima de dicho límite. El número de dichos complejos coincide muy estrechamente con el que existe en los depósitos de loess del centro de Europa, situados también por encima del mismo. Este número concuerda además con el del registro de las oscilaciones climáticas, según datos obtenidos a partir de los sedimentos de los fondos marinos.

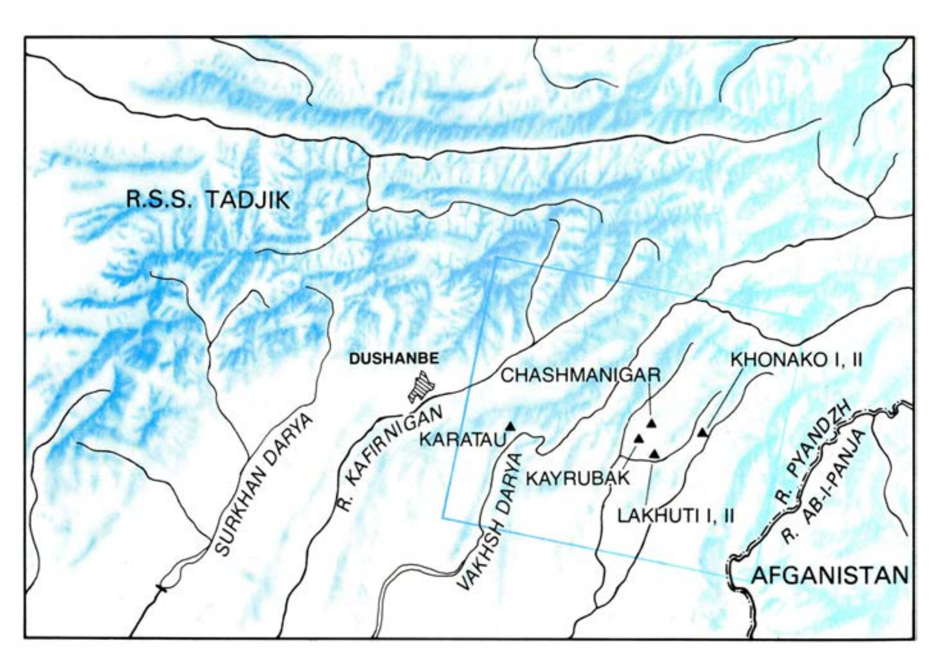
Los análisis polínicos, efectuados bajo la dirección de M. M. Pakhomov, de la Academia de Ciencias del Tadjikistán, indican que, además de las oscilaciones climáticas atribuibles a los avances y retrocesos de las masas de hielo pleistocénico, el clima en esta parte del Tadjikistán se hizo más seco hacia el último millón de años, aproximadamente. A comienzos del Pleistoceno medio predominaban los bosques cubiertos de abedules; existían helechos en las riberas de los ríos y pinos en las partes elevadas. Con el tiempo, las áreas cubiertas fueron disminuyendo gradualmente, siendo sustituidas por praderas, propias de condiciones más áridas. Esta tendencia aún continúa.

Los instrumentos líticos que atestiguan la presencia humana en esta área están siempre asociados con los paleosuelos situados por encima del límite paleomagnético Matuyama-Brunhes. Se han descubierto en el quinto, el sexto y el séptimo complejo edáfico (contando a partir de la superficie). Su posición por encima del límite significa que esta industria tiene menos de 690.000 años. Pero, ¿cuántos menos? Por fortuna, otro dato paleomagnético, el episodio Blake, nos viene en ayuda para acotar esta posición. El episodio Blake, una breve inversión de la polaridad Brunhes, ocurrió hace unos 110.000 años; las pruebas de ello las tenemos, en el loess, por encima del quinto complejo edáfico en tres de las doce secciones donde se descubrió la existencia de utillaje. Lógicamente, hubo de transcurrir cierto lapso de tiempo para las sucesivas acumulaciones de loess y el desarrollo de los suelos situados entre el séptimo y el quinto complejo. Por consiguiente, parece razonable y nada descabellado estimar la edad de los tres complejos edáficos donde fue hallado el utillaje, dentro de un período comprendido entre la parte más reciente del Pleistoceno medio y la más antigua del Pleistoceno superior. En términos de cronología absoluta ello significa un intervalo comprendido entre hace 250.000 y 130.000 años; para la terminología arqueológica del Viejo Mundo ello equivaldría a decir los tiempos finales del Paleolítico inferior.

La datación por termoluminiscencia se basa en el hecho de que un cierto número de sólidos cristalinos, como las partículas de cuarzo del loess del Tadjikistán, acumulan energía de las radiaciones que ionizan el medio a partir del



PRINCIPALES ASENTAMIENTOS DEL PALEOLITICO INFERIOR de Asia. Los triángulos, tales como los que señalan Chukutien en China septentrional y Latamne en Siria, indican que la edad atribuida a la localidad se basa en datos geológicos seguros. Los rectángulos, tales como los que señalan Patjitan en Java y Barda Balka en Irak, indican la ausencia de un contexto geológico válido. De las 26 localidades que figuran en el mapa, sólo nueve corresponden al primer grupo. Dos de estas nueve, Karatau y Lakhuti I, están en la región loéssica de la depresión afgano-tadjik, la zona sometida a estudio por los autores.



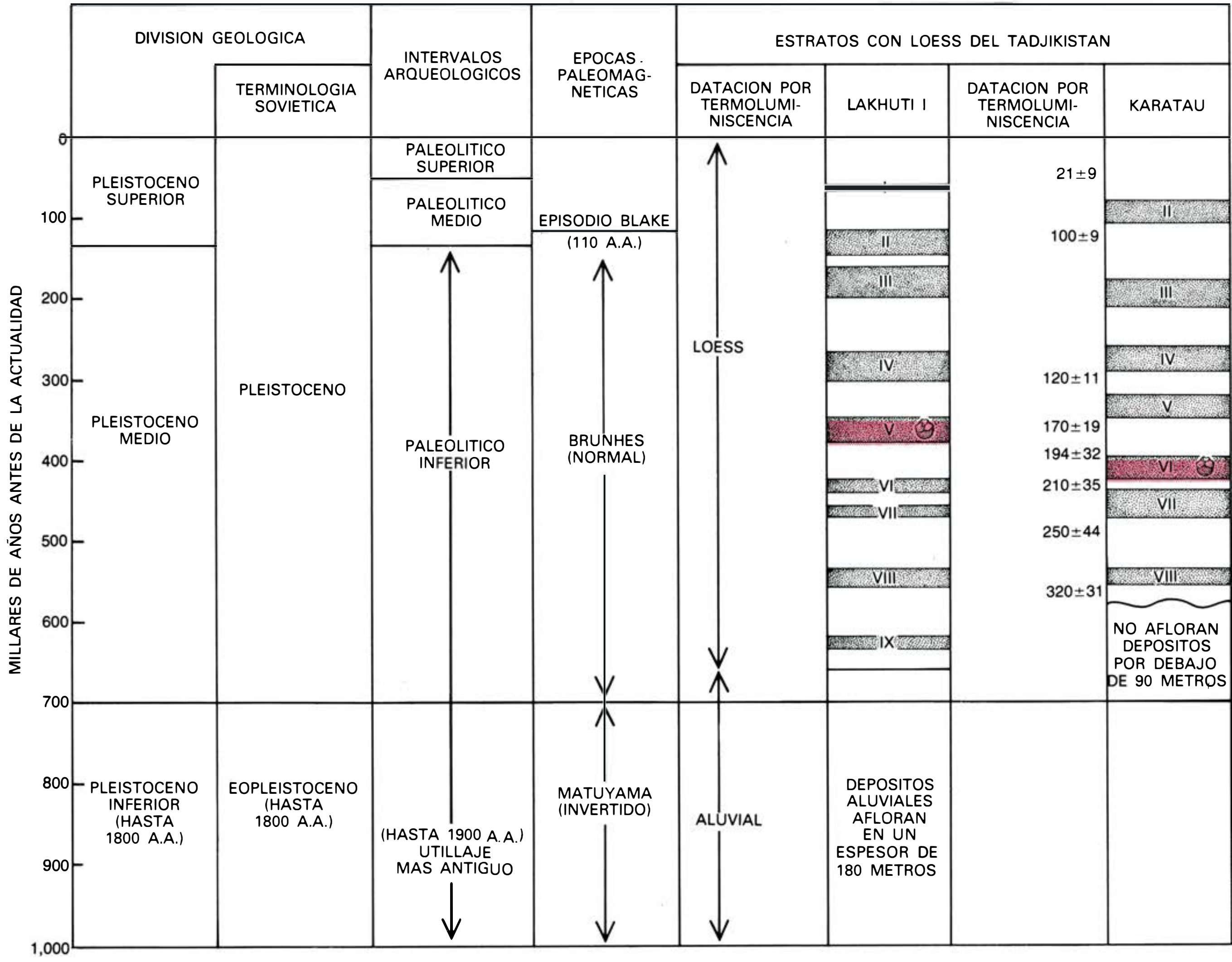
AREA SUDORIENTAL de la República Socialista Soviética del Tadjikistán. Se caracteriza por la existencia de espesos depósitos de loess, que suponen un registro de dos millones de años de cambios en el clima y el ambiente del Pleistoceno. Se señalan cinco localidades (*triángulos*) donde se recuperaron útiles líticos.

momento en que se integraron en los sedimentos; y la devuelven en forma de luz si reciben energía calorífica. La intensidad de la luz emitida es proporcional al tiempo que llevan incorporadas en el material sedimentario. Estos sólidos no acumulan energía mientras permanecen expuestos a las radiaciones ultravioletas solares en la superficie; por así decir, se mantienen a cero hasta su integración sedimentaria.

Un primer intento llevado a cabo por el geólogo ucraniano V. N. Shelkoplyas en la aplicación de la termoluminiscencia para datar las partículas de cuarzo en el loess del Tadjikistán proporcionó al fin un primer punto de apoyo a la estimación de la edad obtenida por el método paleomagnético. Shelkoplyas calculó la edad para cinco de los afloramientos del loess. La más antigua, para loess situados en la actualidad a más de 150 metros por debajo de la superficie, era de 880.000 ± 98.000 años (antes de la actualidad). La más moderna, para loess situados aproximadamente a 10 metros de la superficie, era de 21.000 ± 9000 años. La determinación de edades por termoluminiscencia referidas a la industria lítica (es decir, las dataciones que indican la edad de los depósitos de loess situados por encima o por debajo de los suelos en los que se encontraron los útiles) van de un mínimo de 110.000 ± 13.000 años (loess situado por encima –más joven por tanto– que el suelo número 5 en la localidad Khonako II) a un máximo de 280.000 ± 48.000 años (loess por debajo –más antiguo por ello– que el suelo núm. 7 en Khonako I).

Los resultados de Shelkoplyas son necesariamente aproximados, pero su notable concordancia con los datos paleomagnéticos refuerza su credibilidad. Por ejemplo, la edad por termoluminiscencia de una capa de loess en Kayrubak, situada por el método paleomagnético en el límite Matuyama-Brunhes (hace 690.000 años), es de 650.000 ± 78.000 años, y la edad, por el mismo sistema, de loess de equiparable posición en Lakhuti es de 720.000 ± 82.000. Conviene observar que las dataciones por termoluminiscencia tienden a rebajar la edad de las muestras de cuarzo de manera progresiva con la antigüedad de las mismas. Por ello, puede razonablemente esperarse que la antigüedad de los paleosuelos del Tadjikistán donde se ha encontrado instrumental lítico sea mayor que la sugerida por los datos paleomagnéticos, y de este modo se halle más cerca de la estimación inicial entre los 250.000 y los 130.000 años.

Se excavaron dos de las doce localidades que han proporcionado industria lítica en estos paleosuelos, Lakhuti I y Karatau. El trabajo, inciado en 1973 bajo la dirección de uno de nosotros (Ranov), está todavía en curso. Hasta ahora han sido preparados 276 metros



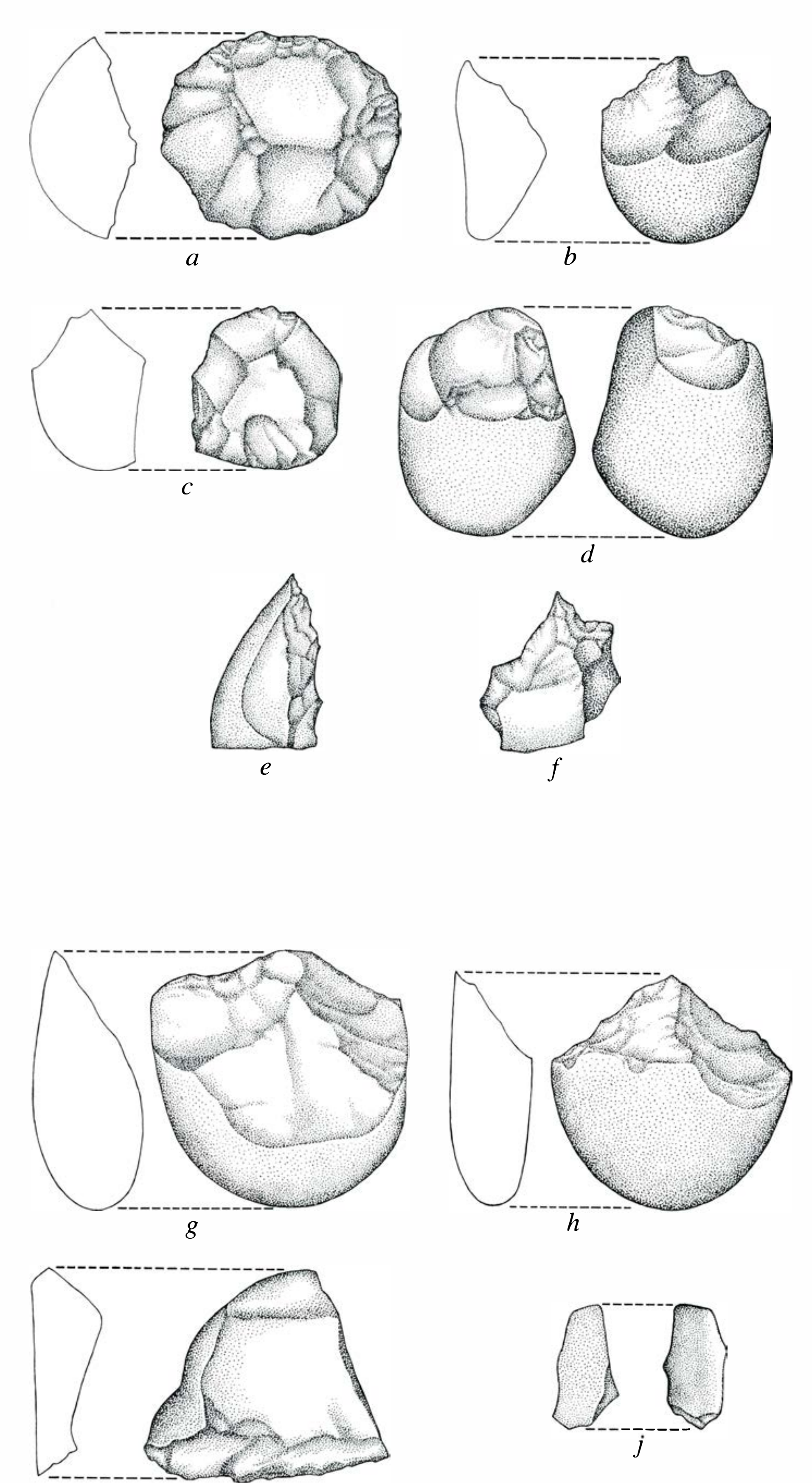
SUELOS FOSILES en las dos series de loess donde se extrajo instrumental lítico (*derecha*). Tales yacimientos aparecen en el diagrama en relación con dos escalas cronológicas distintas. La primera está basada en el paleomagnetismo de las partículas magnéticas del loess. La segunda lo está en la termoluminiscencia de las partículas de cuarzo. Las edades están expresadas en miles de años. Figuran también las divisiones geológicas del Pleistoceno y los intervalos arqueológicos del Paleolítico. Ocho suelos fósiles (*en notación romana*) en Karatau y nueve en Lakhuti I poseen una antigüedad inferior a los 690.000 años. Los instrumentos líticos extraídos proceden del suelo número VI de Karatau y del suelo número V de Lakhuti I (*sombreados en color*).

cuadrados del suelo núm. 5 en Lakhuti I y 124 del suelo núm. 6 en Karatau. En ambas localidades, recuperamos utillaje en un solo horizonte; lo propio sucede en las diez localidades restantes. En los puntos excavados se hallaron en una zona de 1,5 metros de más intensa alteración, situada por encima del horizonte carbonatado. Esta posición indica que los cazadores ocupaban dichas zonas exclusivamente en épocas de clima óptimo, cuando las condiciones de temperatura y pluviosidad eran más favorables tanto para los animales como para las plantas. Los útiles proceden de material metamórfico, que se presentaba en forma de guijarros redondeados por acción de la corriente.

En la localidad de Karatau se han encontrado hasta unas 200 piezas. Son de dos clases principales: guijarros trabajados y lascas. Los guijarros trabajados se obtenían modificando los guijarros naturales de manera sumaria, con arranque de unas cuantas esquirlas de un extremo. Estos sencillos útiles pueden dividirse a su vez en dos categorías: los "choppers" o unifaces y los "chopping tools" o bifaces primitivos. La distinción entre ambos es la siguiente. En un "chopper", el borde cortante se obtenía arrancando esquirlas de una sola cara del guijarro, mientras que un "chopping tool" se preparaba haciendo saltar esquirlas de ambos lados.

La segunda clase de instrumentos, las lascas, parecen haberse obtenido a partir de dos sistemas. El más corriente consistía en efectuar sobre el guijarro un golpe directo que lograba arrancar una única pieza; frecuentemente ello determinaba útiles con sección transversa en forma de cuña, como un gajo de naranja. El otro método consistía en disponer en un extremo del guijarro una "plataforma de percusión" haciendo saltar una o varias esquirlas pequeñas. El golpe subsiguiente, que originaba la lasca, estaba dirigido contra esta plataforma preparada. El 70 por ciento de todas las lascas enteras de Karatau muestran todavía la superficie original del guijarro en la cara por donde fue aplicado al golpe para arrancarlas Esto prueba que, incluso en los casos en que se empleaba una plataforma de percusión, los autores de esta industria obtenían con mayor frecuencia las lascas a partir del guijarro sin más.

El suelo núm. 5 en Lakhuti I es un poco más moderno que el suelo núm. 6 de Karatau. Entre las 450 piezas obtenidas en Lakhuti I, la proporción de lascas con plataforma de percusión es mucho más alta que en Karatau. Se recogieron, además, algu-



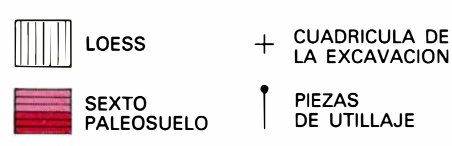
INSTRUMENTOS LITICOS SIMPLES de Lakhuti I (*grupo superior*) y Karatau (*grupo inferior*) dibujados a mitad de su tamaño natural. La diferencia entre los "chopping tools" –bifaces– (*c, d*) y un "chopper" –unifaz– (*b*) consiste en que las esquirlas eran arrancadas en ambas caras de un guijarro para preparar un "chopping tool" y en una sola cara para preparar un "chopper". Las dos lascas apuntadas y retocadas (*e, f*) deben haber sido seguramente empleadas para perforar. Un guijarro (*a*) muestra las facetas originadas por el arranque de varias esquirlas, quizá para ser empleadas ellas mismas como útiles. Las piezas de Karatau comprenden dos "choppers" (*g, h*), una lasca que fue retocada lateralmente (*i*) y una lasca primaria sin retocar (*j*), es decir, material bruto para un instrumento determinado.

nos núcleos de forma discoidal, que indican la costumbre de arrancar varias lascas de un mismo guijarro. La proporción de lascas respecto a choppers (unifaces) y chopping tools (bifaces) es también más alta en Lakhuti I que en Karatau, y parece existir un mayor refinamiento en el retoque secundario de las lascas. Estas diferencias podrían indicar que los ocupantes de la localidad más reciente habían logrado alcanzar un mayor desarrollo técnico en la preparación del instrumental lítico. Sin embargo, cuando se trata de una tecnología tan simple como en ambos casos, resulta difícil interpretar qué cambios son acumulativos y cuáles progresivos.

Aunque en ambas localidades pueden distinguirse los dos tipos de útiles de piedra, choppers (unifaces) y chopping tools (bifaces), las lascas no corresponden a categorías fijas. Sólo podemos describirlas en términos generales como lascas con retoque lateral, lascas con muescas y fragmentos bifaciales, es decir, pequeños útiles originados por arranque de esquirlas, también pequeñas, en ambas caras. (Los últimos son extremadamente raros y las industrias líticas de ambas localidades no pueden interpretarse en realidad como industrias de bifaces típicos, cual si se tratara de las tan conocidas hachas de mano del Chellense y Acheulense del Paleolítico inferior.)

Entre los hallazgos de Lakhuti I se encuentran unos pocos fragmentos de huesos de animales. Fueron identificados como restos de gacelas y de un cérvido. Ungulados afines a éstos pudieron haber sido relativamente abundantes en la estepa abierta, pero aun así la creencia en la existencia de una economía bien establecida sobre la caza y la recolección de vegetales en Asia central durante esta época, fines del Pleistoceno medio, no resulta totalmente convincente. El paisaje de la región comportaría seguramente la existencia, además de la estepa, de un somontano; la diferencia en altura entre las alineaciones montañosas fracturadas y los valles fluviales debía de ser considerable. Hoy, la localidad de Lakhuti I se halla junto al río Obi-Mazar. Karatau, en cambio, se levanta a más de 1000 metros por encima del curso fluvial más próximo, el Vaksh. En conclusión, parece más verosímil que el modelo de actividad humana estuviese basado en la explotación de la región por bandas errantes de cazadores que penetrarían en la depresión afganotadjik durante los períodos interglaciares, climáticamente más favorables para ocupar de manera temporal pequeños campamentos.

¿A qué grupo humano pertenecían esos cazadores y recolectores? Mientras no se descubran restos esqueléticos de ellos, la pregunta no puede contes-



EXCAVACION EN KARATAU, con un corte en el talud del loess por encima del río Vaksh que puso al descubierto una superficie de 124 metros cuadrados. El bloque diagrama muestra la parte sur de la localidad; una fractura (*color oscuro*) desplazó el complejo del sexto suelo a una altura alrededor de tres metros, manteniendo la parte izquierda del suelo a un nivel más bajo que la derecha. El complejo edáfico está constituido por tres elementos: un suelo superior alterado (*color claro*), una gruesa costra de carbonato cálcico (*gris*) y la roca madre del suelo (*color oscuro*) debajo de la costra. Las 25 piezas descubiertas en esta parte de la localidad (*alfileres negros*) aparecieron en el suelo superior. La localidad de Karatau ha suministrado unos 200 "choppers" y lascas. La ilustración fue realizada por A. Tomko.

tarse con base alguna. Se sabe que el homínido más antiguo que habitó Asia pertenecía, con una sola posible excepción, a la especie *Homo erectus,* cuyos restos en Java y China poseen una antigüedad de 1,5 y aun quizá 1,9 millones de años. (La posible excepción es *Meganthropus,* representado por un fragmento de mandíbula encontrado en Java, asignado por algunos al género *Australopithecus.*) ¿Qué poblaciones de *Homo erectus* o de descendientes suyos se desplazaron desde las sabanas de latitud baja hasta los bosques caducifolios del norte para llegar a la estepa semiárida? Por lo que se refiere en particular a la depresión afgano-tadjik, hasta hace poco muchos prehistoriadores sostenían que los hombres primitivos no habían alcanzado estas regiones tan áridas y con un clima tan riguroso del Asia central hasta hace unos 100.000 años, a fines del Paleolítico inferior y comienzos del medio. Esta conclusión se basaba en la creencia de que ninguna población de cazadores podría haberse adaptado a esas condiciones de estepa.

Los hallazgos arqueológicos del Tadjikistán muestran que esta hipótesis no es correcta. Por el contrario, hace unos 250.000 años, si no más, algunos pueblos primitivos cazadores vivían en la depresión afgano-tadjik a tenor de las fluctuaciones experimentadas en la disponibilidad de los escasos recursos. Obviamente tuvieron que adaptarse a las condiciones de los rudos altiplanos semiáridos. Que se sepa, en ningún otro punto del Viejo Mundo han existido similares adaptaciones culturales.

Debe admitirse, pues, que la competencia adaptativa de estos pueblos cazadores era mayor de lo que generalmente se ha venido suponiendo. En realidad puede augurarse que se recogerán pruebas de la existencia de pueblos cazadores del Paleolítico inferior en muchos puntos de Asia donde no se esperaba encontrarlas. Así, por ejemplo, las referencias sobre hallazgos de hachas de mano bifaces del Kazakhstán central podrían explicarse como reflejo de una mayor penetración hacia el norte por estos cazadores cuya presencia se había reconocido antes. Por su parte, los hallazgos de Korolevo, en los Cárpatos de Ucrania, por V. N. Gladilin, del Museo Arqueológico de Kiev, muestran también que los cazadores de fines del Pleistoceno medio podían sobrevivir a los severos y prolongados inviernos. Con este hecho no habían contado los especialistas que se ocupan de esta época.

Los instrumentos líticos hallados en Lakhuti I y en Karatau son muy simples: algunos prehistoriadores podrían considerarlos incluso toscos o primitivos. Su presencia en Tadjikistán hace unos 250.000 años constituye, no obstante, una prueba incontrovertible de que una sencilla asociación instrumental de núcleos y lascas bastaba para satisfacer los requisitos adaptativos de una demanda de hábitat. De manera clara, los útiles muestran que eran necesarios –en términos de percutir, cortar, raer, perforar y otras operaciones– para la supervivencia de los cazadores del Asia central bajo tales condiciones. La conclusión en este caso es que una tecnología elemental no implica de manera necesaria un potencial limitado en adaptaciones.

Si, como parece, el hombre primitivo habitó la depresión afgano-tadjik especialmente durante las épocas interglaciares, entonces no sólo esta región sino también otras áreas extensas del Asia pudieron haber sido testigos de movimientos de poblaciones humanas instados por las oscilaciones climáticas de larga duración. Tales movimientos consiguieron impedir dilatados aislamientos entre los pueblos de las regiones asiáticas. Si las cosas sucedieron así, ello estaría en desacuerdo con la opinión de lo que fue Asia durante el Paleolítico, opinión que sigue contando con no pocos adeptos. Este punto de vista sostiene que en Asia coexistieron durante mucho tiempo, en el Pleistoceno medio y superior, y por separado, dos tecnologías líticas distintas, propias de dos tipos de poblaciones que vivían aisladas: la industria de los "choppers" y los "chopping tools" y la de las hachas de mano bifaces.

La idea corrientemente aceptada es la de que la tradición de los "choppers" y "chopping tools" procede de la tecnología heredada de los primitivos homínidos que desde Africa se dispersaron por todo el Antiguo Continente. Esta tradición se caracterizaba por el uso de instrumentos líticos relativamente simples con un mínimo de retoque. En cambio, para la tradición del hacha de mano, que suele reputarse más avanzada, se admitía también que había surgido de Africa y Europa, pero que no había penetrado más allá del subcontinente indio. Se creía que la tradición de los "choppers" y "chopping tools" se había mantenido aislada en Asia: una especie de remanso cultural confinado al sur del Himalaya, sudesde de Asia y Lejano Oriente, que contribuyó escasamente a las grandes corrientes del desarrollo humano.

La visión convencional presupone la existencia de mecanismos de aislamiento cultura o natural, que habrían mantenido separadas ambas tradiciones durante centenares o milenios de años. Por otra parte contempla los modelos de tecnología de instrumental lítico como reflejo de dos tradiciones culturales diferentes. Hasta cierto punto, una hipótesis de dos culturas apoya la idea de que la disyunción entre los pueblos del este y los pueblos del oeste tendría un origen muy antiguo, situado en algún momento del Pleistoceno inferior.

Los descubrimientos del Tadjikistán y otros recientes hallazgos muestran una superposición geográfica de las dos tradiciones de utillaje, lo cual desbarata la hipótesis convencional. La nueva visión que sustituye a la de las dos culturas descarta la posibilidad de un aislamiento regional de larga duración en Asia, a la vez que reconoce que, por lo que atañe a la tecnología lítica, existía una considerable variabilidad y una diversidad regional en cualquier punto de Asia de latitud media y baja. Esta diversidad podría responder a varias causas. Entre ellas pueden ser tomadas en consideración: las respuestas adaptativas de corta duración a los cambios ambientales locales, las necesidades funcionales referidas a trabajos estacionales y actividades específicas similares, la "deriva" cultural y hasta los movimientos migratorios. Al mismo tiempo, un flujo significativo de genes entre poblaciones humanas en contacto contribuiría a evitar importantes aislamientos genéticos de cierta duración.

Parece, pues, que pretender asignar los útiles del Paleolítico inferior del Tadjikistán a alguno de los componentes de la insostenible hipótesis de las dos culturas constituye una inútil pérdida de tiempo. La industria de núcleos (guijarros trabajados) y lascas del Tadjikistán debe valorarse, no a partir de amplias generalizaciones, sino sobre bases más regionales; y han de investigarse localmente las influencias que determinan sus características. Esto queda claro cualesquiera que sean las otras posibles sorpresas que pueda depararnos el estudio del loess del Asia central. En esta área podría muy bien existir un registro de la actividad humana desde comienzos del Pleistoceno medio, registro que quizá proporcionaría la clave para la comprensión de la evolución biológica y cultural de nuestros antepasados del Paleolítico en Asia.

La conducta adictiva

Sea la adicción a las drogas, al alcohol o al tabaco, el problema reside en la conducta autodestructiva que genera. Un tratamiento de mantenimiento con metadona ha logrado devolver la salud a muchos adictos a la heroína

Vincent P. Dole

Los graves problemas sociales y personales que se derivan de la adicción, sobre todo a las drogas, pero también al alcohol y al tabaco, son extraordinariamente difíciles de afrontar. El índice de fracasos de los programas de "desintoxicación" y rehabilitación es alto. Un programa que da buenos resultados, siempre que se aplique correctamente, es el mantenimiento de adictos a la heroína con metadona, un análogo de la heroína. El programa de mantenimiento con metadona realizado en la ciudad de Nueva York, en el que he participado como médico y administrador desde su inicio, constituye un ejemplo de la eficacia de estos programas. Nos proporciona, además, una base para reflexionar sobre la adicción y los problemas que plantea su tratamiento.

Nadie sabe qué es lo que realmente ocurre con la conducta adictiva. Ni animales ni niños nacen con ansia por los narcóticos, los sedantes, el alcohol, la cafeína, el tabaco, las anfetaminas o la cocaína. Mucha gente, incluso personas que más tarde se convertirán en toxicómanos, alcohólicos o fumadores, reaccionan con náuseas y mareos a la primera toma. Resulta extraño, cuando menos, que algunos no sólo logren tolerar esos tóxicos, sino que desarrollen una necesidad irresistible de ellos.

Abundan las teorías. La sabiduría popular dice que los adictos toman drogas por placer; afirmación esta veraz, aunque incompleta. ¿Cómo logran esas sustancias, en un comienzo repugnantes, hacerse placenteras? ¿Por qué sólo algunos se tornan adictos siendo tan amplia la población expuesta? ¿Quizá la minoría adicta reacciona de manera distinta ante las drogas, o es que la mayoría no adicta tiene menos necesidad de gratificación?

Los tratados médicos antiguos atribuían la adicción a una personalidad adictiva, es decir, a un defecto del carácter diagnosticado a partir del hecho de que el sujeto era adicto. Estas explicaciones circulares poco aclaraban sobre las razones de la especificidad de cada adicción. Los opiómanos anhelan opiáceos, los alcohólicos, alcohol y los fumadores, cigarrillos. Incluso en personas con adicciones múltiples se conservan anhelos por sustancias específicas. ¿Por qué esto es así?

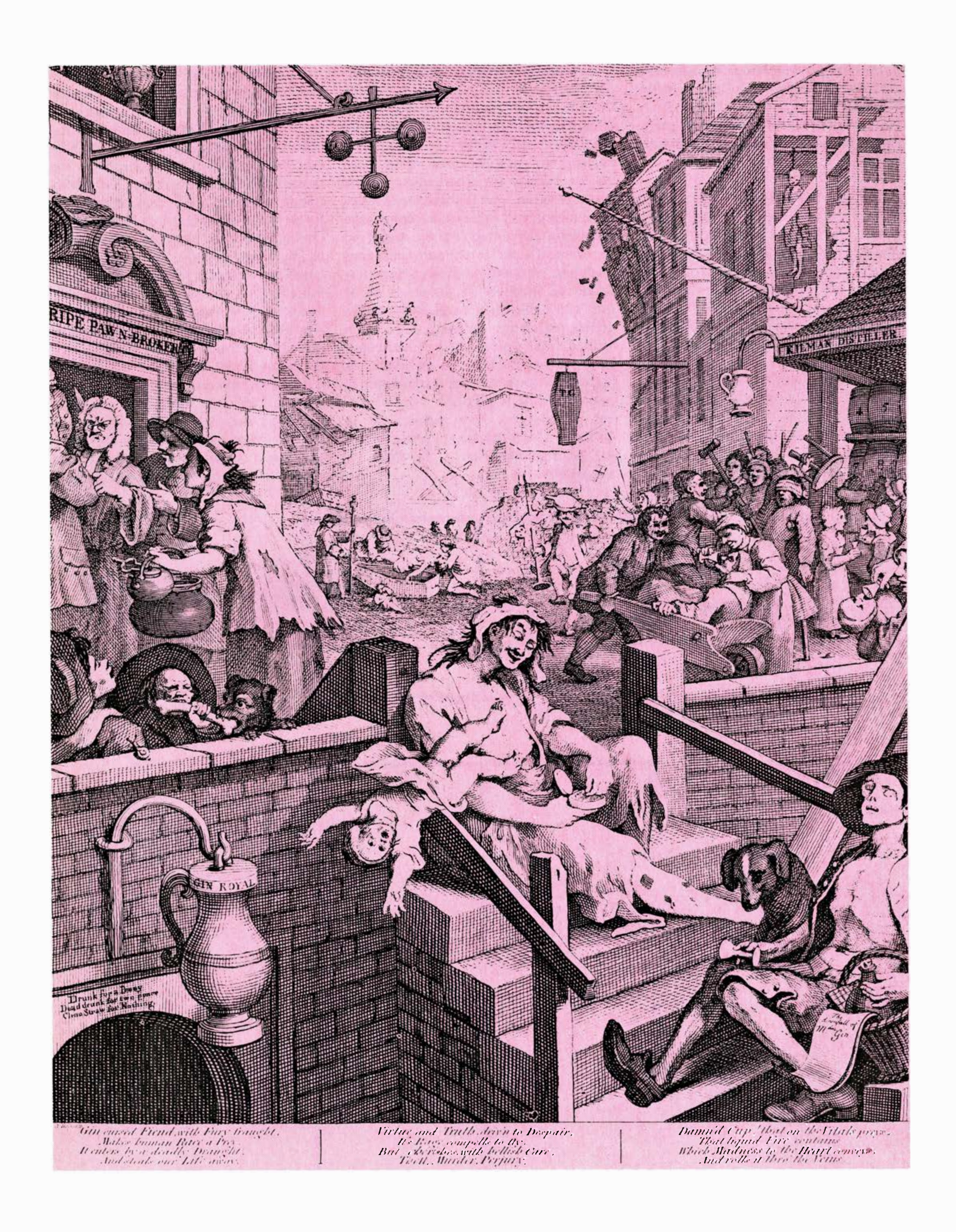
Compitiendo con la ineficacia de las explicaciones psicológicas encontramos teorías que manejan argumentos puramente farmacológicos. Una teoría hipersimplificada dibujaría al adicto como una persona atrapada en el hábito de consumir drogas para eludir el castigo: cuando desaparece el efecto de una droga adictiva, aparecen los síntomas de abstinencia, que empujan al adicto a buscar alivio en otra dosis. Esta explicación resultó verosímil y aclaraba la recurrente desesperación de la conducta de búsqueda del tóxico, pero no pudo explicar una cuestión fundamental: ¿por qué la mayoría de los adictos recaían después de haber superado el período de necesidad imperiosa del producto?

El tratamiento más antiguo de la adicción es la desintoxicación (término procedente de la refutada teoría de que los adictos padecen una acumulación de toxinas en su organismo). En este tratamiento se suprime la droga bajo vigilancia médica. En el período de abstinencia, los síntomas difieren según la droga de que se trate, la magnitud de su uso hasta ese momento y el tipo de asistencia que se reciba en el momento de la supresión; aunque con el control adecuado es posible la desintoxicación de cualquier paciente. A partir de entonces, la abstinencia es posible sin un malestar excesivo. Para emplear una frase técnica, se ha conseguido eliminar la dependencia física.

Algunos investigadores han equiparado adicción y dependencia física, pero esta óptica ignora la parte más difícil del problema: la conducta de búsqueda del tóxico. La dependencia física es simplemente una consecuencia adaptativa ante la ingesta repetida de algunos productos químicos. Muchos fármacos administrados en la práctica médica –por ejemplo, los esteroides– provocan una dependencia física en el sentido farmacológico, sin que aparezca un deseo por la sustancia. De manera semejante, la administración obligada de narcóticos o alcohol a animales que no lo solicitan los hace depender físicamente de ellos, sin que aparezca una conducta adictiva. La gente que toma esteroides y los animales que reciben narcóticos o alcohol suprimen o reducen sustancialmente su consumo cuando la ingesta es optativa.

La distinción entre dependencia y adicción queda demostrada más claramente con el habitual fracaso de las curas de desintoxicación a que se someten los adictos. Si este tratamiento terminase con el hábito, la adicción sería un problema médico sencillo, pero resulta que la mayoría de los consumidores recaen después de haberse librado de su dependencia física. En un intento por reconciliar este hecho con la teoría farmacológica, algunos expertos han invocado el concepto de dependencia psicológica, término que se basa en la idea de que la desintoxicación ha eliminado los determinantes fisiológicos de la conducta adictiva, sin modificar la necesidad psicológica. Como en la teoría de la personalidad adictiva, el concepto de dependencia psicológica se caracteriza por ser una inferencia de muchas conductas que quedan sin explicar.

Una teoría psicofarmacológica alternativa identificaba las recaídas con un reflejo condicionado. Abraham Wikler, trabajando en el Public Health Service Hospital de Lexington, Kentucky, observó que algunos adictos dados de alta de la institución tras haber sido tratados padecían un intenso anhelo cuando volvían a su casa o a su barrio. Pudo descubrir ciertos estí-



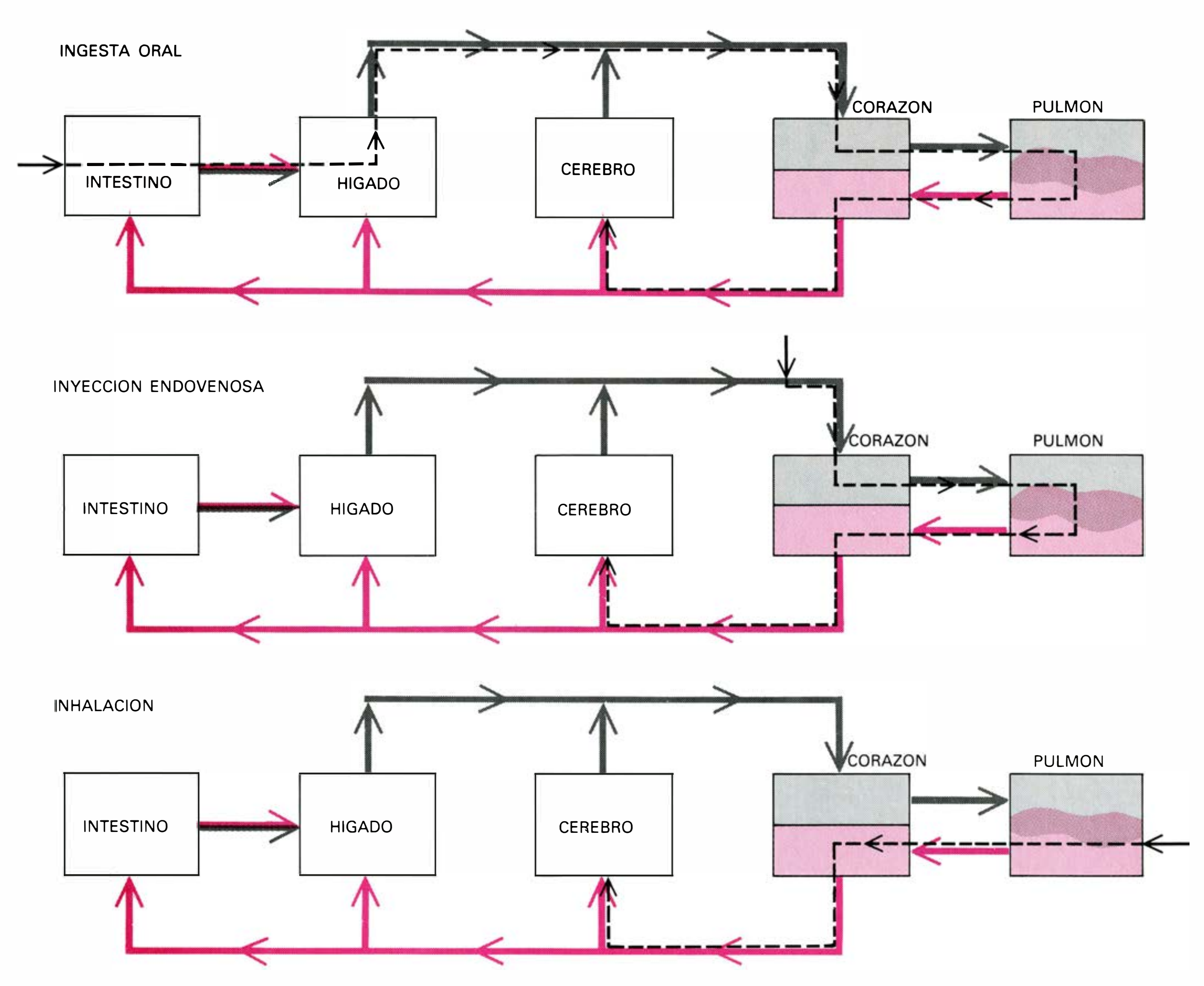
ADICCION AL ALCOHOL simbolizada en el grabado *Gin Lane*, de William Hogarth. La escena representa de forma algo exagerada la epidemia de alcoholismo que se extendió por Inglaterra hace unos 250 años, cuando se podía conseguir alcohol concentrado y barato bajo la forma de "ginebra" (gin). En las escaleras, una mujer borracha deja que su hijo se despeñe. Más abajo, aparece un alcohólico en agonía; a la izquierda, la gente empeña objetos para sacar dinero con qué beber y, al fondo, varias personas se emborrachan o sufren las secuelas del etilismo. Es la ruina de la sociedad.

mulos (códigos) ambientales asociados a la toma de heroína, a los que consideró percutores para reactivar el anhelo. En estas circunstancias podían reaparecer incluso síntomas físicos de abstinencia aguda.

Esta observación permitió albergar la esperanza de que el ciclo adictivo pudiera romperse con un aprendizaje que disociara los códigos de los efectos gratificantes de la droga. Por desgracia, los intentos de aplicación de este principio no han tenido mucho éxito. Se dispone de potentes fármacos capaces de bloquear la acción de los narcóticos, pero su prescripción no ha permitido extinguir la conducta de búsqueda del tóxico en un considerable número de adictos. Aunque los resultados negativos no invaliden necesariamente la teoría, su utilidad práctica parece más que dudosa.

Una teoría sociopsicológica más amplia, que incorpora algunos elementos de la teoría del condicionamiento, está ganando popularidad. Al advertir que la adicción al alcohol y a las drogas está alcanzando proporciones epidémicas en los barrios superpoblados, mucha gente sensible a los fenómenos sociales ha interpretado la adicción como una estrategia para huir de un ambiente opresivo. Desde este punto de vista, cualquier tratamiento médico es superficial, en el mejor de los casos no es más que paliativo y en el peor resulta contraproducente respecto a la tarea fundamental de reformar la sociedad. El atractivo político de la teoría es poderoso, puesto que homologa la prevención de la adicción a los objetivos prioritarios de supresión de la pobreza, el desempleo, el hacinamiento, la discriminación y la falta de instrucción.

Al margen de las dificultades prácticas que entraña prevenir la adicción mediante la abolición de la pobreza, la teoría sociopsicológica no proporciona tratamiento alguno a los que ya son adictos ni explica las diferencias de susceptibilidad individual. La gente tiende a evitar la adicción, incluso en los peores ambientes y, de hecho, es la disponibilidad de la droga, más que la calidad de la vida, lo que parece constituir el factor principal que determina la prevalencia del consumo. La heroína se ha extendido a suburbios y pequeñas ciudades que se suponían inmunes al fenómeno de la adicción; los efectos del alcohol y el tabaco, universalmente disponibles, alcanzan a todos los estratos sociales. Un ámbito que haya escapado de la epidemia de los narcóticos puede que deba su buena suerte a la escasez de la droga más que a una estructura social privilegiada.



RUTAS SEGUIDAS POR LAS DROGAS hasta llegar al cerebro desde el torrente circulatorio; dependen de que el consumo sea por ingestión oral, inyección intravenosa o inhalación. Si drogas como la heroína se inyectan o inhalan, alcanzan el cerebro a altas concentraciones tras una demora de sólo unos segundos. El alcohol también llega a la circulación rápidamente si la ingesta es abundante, pues se absorbe con rapidez en el estómago y los intestinos; atraviesa el hígado sin disminuir notablemente su concentración. Sin embargo, si la metadona se toma por vía oral se absorbe gradualmente en el tramo intestinal y su circulación por la sangre se retarda debido a la captación hepática. Puesto que la absorción de metadona por el hígado y otros tejidos es reversible, su concentración en el torrente sanguíneo permanece bastante uniforme, si se administra una dosis diaria del fármaco.

La mayoría de las autoridades encargadas de controlar la adicción están de acuerdo en la necesidad de recurrir a una estrategia omnicomprensiva: por eso limitan legalmente la disponibilidad de sustancias peligrosas, ponen en marcha medidas para prevenir la adicción –mejorando las condiciones sociales, la educación y el empleo– y reducen la demanda tratando a la gente que ya es adicta. Mientras tanto, cualquiera que examine la confusión de programas que intentan enfrentarse a la epidemia de drogas advertirá la carencia de un conocimiento profundo de lo que es el proceso adictivo. Como ocurría con las epidemias de la Edad Media, la política gubernamental de nuestros días parece contradictoria, ineficaz y determinada por posiciones de partido. Es probable que la situación no mejore hasta que se aclaren los factores biológicos subyacentes a la adicción.

Merece la pena destacar que los animales no pueden llegar a ser adictos. Los investigadores que han utilizado modelos animales para explicar la adicción, o han fracasado, o han tenido éxito bajo condiciones tan especiales que se duda de la utilidad que puedan tener los resultados para resolver el problema humano. Si bien los efectos farmacológicos de las sustancias adictivas que se inyectan en los animales resultan bastante similares a los producidos en el hombre, los animales tienden a evitar esos productos si se les da opción a ello. Se han definido bien los efectos tóxicos de diversas drogas, pero sigue sin respuesta la cuestión fundamental de por qué los adictos humanos las consumen repetidamente. Como pretenden las teorías psicológicas, quizá la ingesta autodestructiva de productos adictivos sea una peculiaridad humana dependiente de procesos cognitivos ausentes en los animales. O quizá no se ha descubierto la combinación ideal de animal experimental, trastornos metabólicos y tensión ambiental. La cuestión es atractiva y sólo puede responderse investigando más.

Como a menudo ocurre en el estudio de las enfermedades humanas, el trabajo clínico con problemas y conductas adictivas ha conducido a hallazgos de importancia general mientras la cuestión fundamental sigue sin resolverse. El estudio de los opiáceos en diversos laboratorios puso de relieve la presencia en varios tejidos de sitios de ligamiento específico; los estudios subsiguientes identificaron los ligandos naturales de estos sitios como polipéptidos, cortas cadenas de aminoácidos. Se sabe ya que el significado de los receptores y de los polipéptidos con afinidad por ellos trasciende su acción narcótica. Se piensa que pudieran estar relacionados con diversas enfermedades crónicas y se especula con el papel que puedan desempeñar en trastornos que van desde la obesidad hasta la esquizofrenia. Sin embargo, hasta donde nos es posible saber, la conducta adictiva no está causada por un exceso o una deficiencia de los sitios de ligamiento o de sus ligandos. No hay pruebas de que sean anormales en las personas susceptibles antes de la adicción o en los adictos desintoxicados antes de su recaída. Tampoco hay ninguna prueba firme que los relacione con el alcoholismo o cualquier otra forma de adicción. Por tanto, la investigación bioquímica no proporciona de momento ninguna orientación teórica para el tratamiento de los adictos. Como en otras enfermedades crónicas de causa desconocida, la adicción debe tratarse sin que comprendamos sus procesos fundamentales. La única base para la decisión es el análisis de la experiencia clínica.

Este fue el punto de vista de Marie Nyswander y del autor cuando iniciaron su trabajo conjunto sobre la adicción a los narcóticos hace 17 años. Entonces, como ahora, las teorías proliferaban, aunque su valor práctico era escaso. Los métodos tradicionales de tratamiento –desintoxicación, psicoterapia y estímulos aversivos (castigo)– fracasaban en miles de adictos a la heroína. El Public Health Service Hospital de Lexington, considerado pionero en los programas de tratamiento, sólo proporcionaba internamiento o psicoterapia. El índice de recaídas en los adictos que se reintegraban a la comunidad era del 90 por ciento. En la siguiente década, los estados de California y de Nueva York reprodujeron el fracaso de la reclusión y la psicoterapia, incluso a mayor escala. Se hacía imperioso el ensayo empírico de otros métodos de tratamiento y la obtención de estadísticas fiables.

Afortunadamente, el hospital de Lexington había atraído a investigadores clínicos de capacidad poco usual. Recurriendo a una población de presos adictos realizaron detallados estudios de los efectos a corto plazo de las drogas adictivas. Sus observaciones proporcionaron una rigurosa clasificación farmacológica, así como definiciones precisas de los conceptos de tolerancia y dependencia física.

Nos fueron de gran utilidad para el planteamiento de nuestra estrategia las diferencias entre los efectos de distintos narcóticos que había descrito el grupo de Lexington. Aunque los opiáceos son farmacológicamente afines, presentan una amplia gama de estructuras químicas con diferentes significados en cuanto al sitio y al tiempo de acción. Esta variación abría la interesante posibilidad de que los adictos a la heroína se mantuvieran mejor tomando un narcótico más benigno. A diferencia del alcohol, que es el mismo en la ginebra que en el whisky, los narcóticos difieren en cuanto a absorción, metabolismo, efectos subjetivos y peligros. Supusimos que los síntomas conductuales de la adicción podrían depender también del narcótico particular que se consumiese.

Esta cuestión no se sometió a estudio, pues la prisión no es el mejor lugar para enjuiciar la conducta. El buen comportamiento en la cárcel no prueba que se haya producido una rehabilitación ni tampoco la alienación de un prisionero constituye una prueba fiable de que sea un psicópata. Para realizar un correcto ensayo de comportamiento social es preciso situar al sujeto en su comunidad, cuyo seguimiento debe prolongarse más allá de los cinco meses de estancia en un hospital carcelario.

En consecuencia, empezamos a hacer comparaciones sistemáticas de los distintos narcóticos que se administraban a heroinómanos crónicos, atendidos en la sala de metabolismo del Hospital de la Universidad Rockefeller. Se acordó que los sujetos pudieran suspender el estudio en cualquier momento. A medida que progresaba el tratamiento se les permitía salir a visitar a sus amistades o a trabajar media jornada. Para detectar la ingesta de drogas no permitidas en el exterior del hospital se procedía a un frecuente análisis de muestras de orina. La mayor parte del tiempo se empleó en reuniones y charlas no dirigidas, predominando la escucha a los pacientes. Fue una suerte para nosotros contar en este estudio con Mary Jeanne Kreek, joven y competente investigadora de farmacología clínica. Participó en los primeros trabajos y continuó como investigadora autónoma. En los últimos años, sus estudios sobre las diferencias metabólicas de los narcóticos han servido para clarificar las importantes diferencias existentes entre la metadona y otros opiáceos.

Nuestros primeros ensayos de mantenimiento utilizando diversos opiáceos (morfina, heroína, oximorfona y otros) fueron descorazonadores. Como ya habían predicho los expertos de Lexington, los pacientes no aceptaban el tratamiento más que unos días. Se volvían irritables entre dosis, pedían

más cantidad de droga y se preocupaban de forma obsesiva por la pauta de administración. Bastaba demorar cinco minutos una dosis para que ya hubiese verdaderos altercados. Realmente, los pacientes no se transformaron en candidatos para la reinserción social.

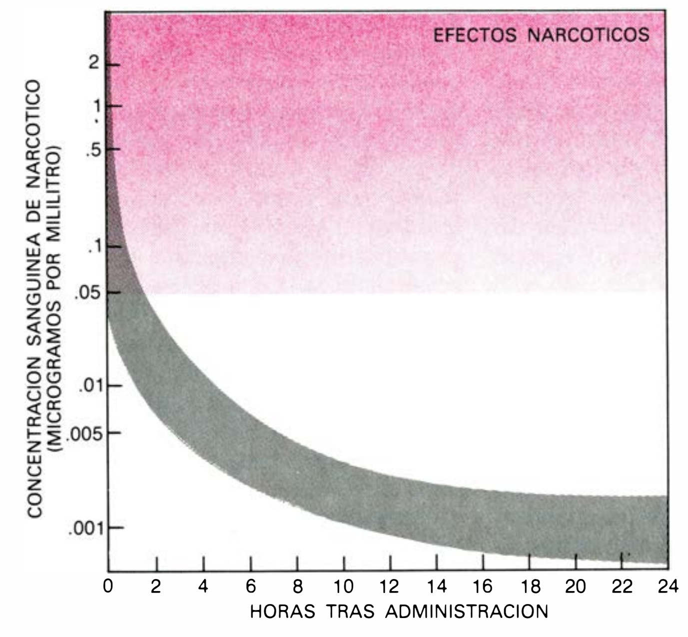
La actitud cambió cuando los pacientes entraron en la fase de estudio con metadona. No esperábamos este efecto y tampoco disponíamos de teoría alguna que lo explicase, pero, partiendo de la ignorancia, aumentamos gradualmente la dosis de metadona hasta alcanzar un nivel constante, relativamente alto: de 80 a 120 miligramos orales, una vez al día. En estas condiciones, los pacientes parecían médica y funcionalmente normales. La medicación no los volvía eufóricos, como ocurría con la morfina y la heroína, y pasaron pruebas de vigilancia y coordinación diseñadas para los pilotos de la Armada estadounidense. (Los tests fueron aplicados por Norman Gordon, de la Universidad Yeshiva.) Y, más importante aún, su anhelo por los narcóticos disminuyó. Iniciaron cursos por correspondencia y tomaron empleos a horas.

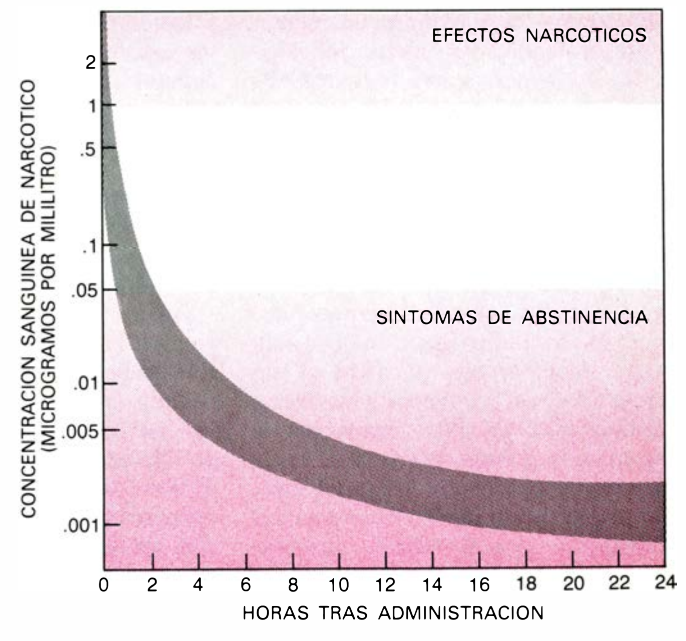
A medida que avanzaba el estudio, en los ocho meses siguientes se permitió a los adictos que utilizaran el centro como residencia, permaneciendo en él mientras era necesario para el programa de administración de la droga, en tanto que podían abandonarlo a su antojo en horas diurnas. A partir de los tres primeros meses, el abuso en el consumo de narcóticos fue relativamente pequeño. La mayoría de los pacientes intentaron recurrir a la heroína en sus primeras salidas del hospital y comprobaron que era ineficaz, puesto que la tolerancia inducida por la dosis diaria de metadona bloqueaba los habituales efectos de la droga. Se efectuaron ensayos sistemáticos en los que se inyectaban determinadas dosis de heroína y de otros narcóticos por vía endovenosa, comprobándose que esta tolerancia bloqueaba los efectos de todos los opiáceos. En los pacientes mantenidos con una amplia dosis diaria de metadona no se producía el "flash" (subida) de los opiáceos.

Se trataba, pues, de un fenómeno completamente imprevisible: la administración diaria de un potente opiáceo no narcotizaba a los adictos. La tolerancia específica de éstos era lo bastante alta como para eliminar los efectos narcóticos, haciendo desaparecer las ansias de heroína. Por primera vez en muchos años, eran capaces de conducirse como ciudadanos normales y volver a sus domicilios sin mezclarse con drogas. Parecía, por fin, que estas personas eran ya hábiles para una rehabilitación social.

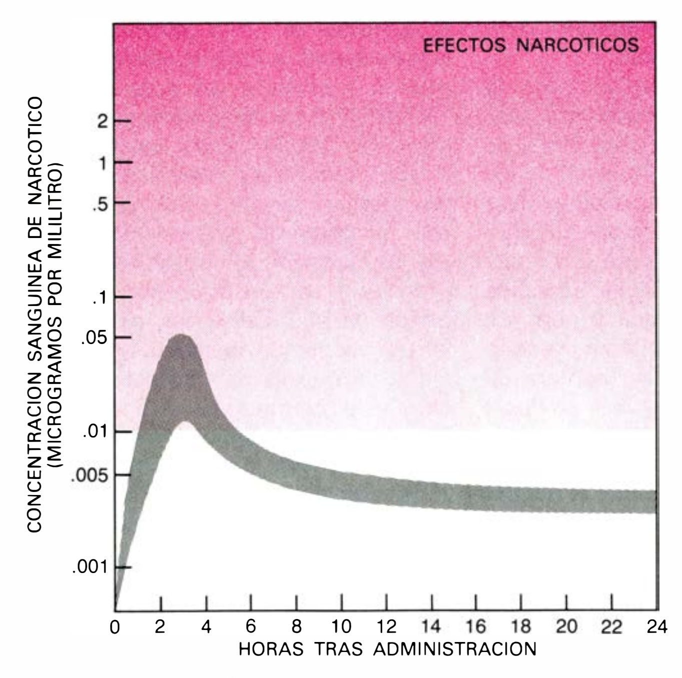
En consecuencia, el trabajo se dividió en dos ramas: (1) el desarrollo de servicios de rehabilitación para adictos estabilizados desde el punto de vista médico y (2) la continuación en las investigaciones sobre farmacología básica. Los programas terapéuticos, que no voy a discutir en detalle, han logrado que, en los últimos 15 años, miles de adictos hasta entonces intratables lleven una vida de ciudadanos normales. El lector interesado en el asunto encontrará numerosas publicaciones donde se trata de los servicios globales que deben acompañar a la terapia con metadona. Con esta ayuda y un mantenimiento farmacológico efectivo, los pacientes presentan un extraordinario potencial para la rehabilitación. Todos los que se admitieron en el estudio inicial tenían historiales delictivos, carecían de trabajo estable, vivían sin hogar ni familia fija y habían perdido completamente el respeto por sí mismos. Hoy, la mayoría lleva una vida responsable, mantienen a su familia y cuidan del hogar. Muchos de ellos siguen tomando metadona.

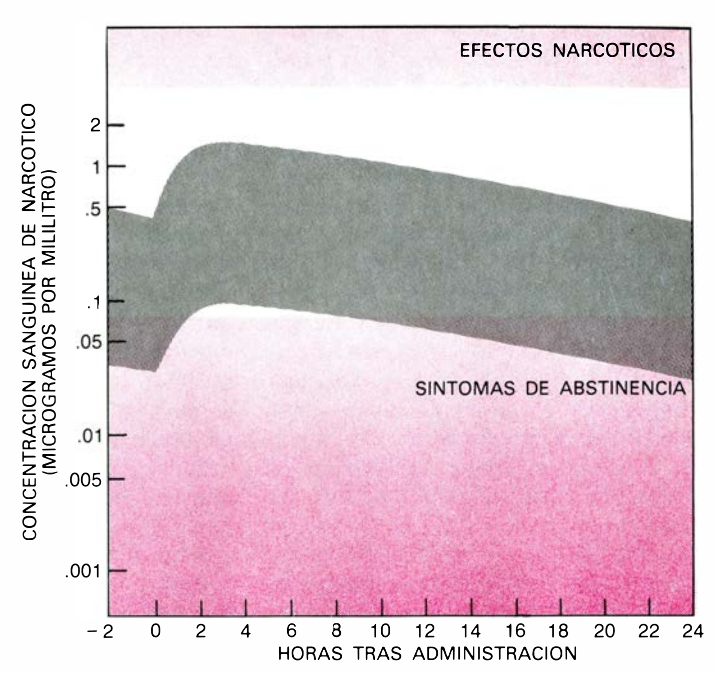
Desde una perspectiva teórica nos preguntábamos cómo un potente narcótico podía eliminar el anhelo sin provocar efectos indeseables. Algunos años más tarde encontramos la respuesta: el fenómeno dependía de la estabilidad de la concentración del fármaco en sangre. A diferencia de la heroína y de otras drogas de acción corta, la metadona (cuando se administra oralmente en dosis adecuadas para estabilizar al adicto) mantiene sus concentraciones a un nivel constante durante unas 24 horas. Deben satisfacerse dos condiciones: la dosis diaria de metadona ha de ser adecuada y la tolerancia del sujeto tiene que ser lo bastante alta como para bloquear el efecto narcótico de la dosis. Si la farmacología del producto se conoce bien, ambas condiciones pueden cum-

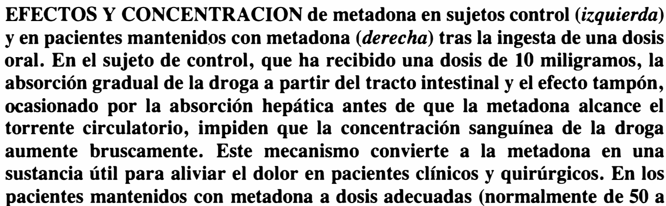




CONCENTRACION DE MORFINA en sangre y sus efectos funcionales tras la inyección endovenosa de 10 miligramos en un sujeto normal (*izquierda*) y en un adicto (*derecha*). Aunque la concentración (*gris*) viene a ser la misma en ambos casos, los efectos funcionales difieren completamente debido a la tolerancia y a la dependencia física del adicto. El período transcurrido entre la desaparición del efecto narcótico y el inicio de los síntomas de abstinencia es corto en el adicto y éste siente pronto la necesidad de una nueva dosis. El organismo produce rápidamente morfina a partir de la heroína inyectada.







EFECTOS Y CONCENTRACION de metadona en sujetos control (*izquierda*) y en pacientes mantenidos con metadona (*derecha*) tras la ingesta de una dosis oral. En el sujeto de control, que ha recibido una dosis de 10 miligramos, la absorción gradual de la droga a partir del tracto intestinal y el efecto tampón, ocasionado por la absorción hepática antes de que la metadona alcance el torrente circulatorio, impiden que la concentración sanguínea de la droga aumente bruscamente. Este mecanismo convierte a la metadona en una sustancia útil para aliviar el dolor en pacientes clínicos y quirúrgicos. En los pacientes mantenidos con metadona a dosis adecuadas (normalmente de 50 a 100 miligramos), la concentración de la droga en sangre se mantiene siempre por encima del umbral de los síntomas de abstinencia. (En algunos casos puede necesitarse una dosis más alta si la eliminación de metadona es anormalmente rápida.) La concentración máxima en un paciente estabilizado se mantiene por debajo del umbral de aparición de los efectos narcóticos. Cuando un paciente en tratamiento recibe una dosis diaria correcta de metadona, se comporta funcionalmente de forma normal, está protegido de los efectos narcóticos por su tolerancia farmacológica a la droga y evita los síntomas de abstinencia por la constante presencia de la metadona en sangre.

plirse sin dificultad. Empezando con una cantidad pequeña, y aumentando gradualmente la toma diaria a medida que aparece la tolerancia durante un período de cuatro a seis semanas, un médico competente puede estabilizar a un adicto con metadona con la misma facilidad con que se mantiene a un cardíaco con una dosis constante de digital.

La farmacología merece un examen más pormenorizado. La metadona pierde su poder sedante y analgésico cuando se administra repetidamente a dosis fijas. El sujeto se hace tolerante y, si la dosis se aumenta gradualmente hasta un nivel estable de mantenimiento, los efectos nunca son importantes, puesto que esta tolerancia específica a los narcóticos crece también con la dosis.

Al mismo tiempo, el sujeto se va adaptando a la presencia del narcótico en su sistema nervioso y necesita mantener su concentración por encima de un mínimo para evitar los síntomas de abstinencia: se ha provocado la dependencia física. Entre los límites que separan la narcosis de la abstinencia hay un espacio funcional. Si la concentración de metadona circulante se mantiene por debajo del nivel de los efectos narcóticos y por encima del umbral de los síntomas de abstinencia, el sujeto estará despejado y se sentirá bien. Lo que el estudio con metadona ha proporcionado es la posibilidad de que el paciente mantenga su ánimo normal mediante medicación; ello resultaba imposible con la administración de heroína, morfina o de narcóticos similares. El estudio que se llevó a cabo posteriormente explicó el por qué.

Cuando un adicto se inyecta una solución concentrada de heroína en vena, el primer fenómeno se concreta en un acusado impacto de la droga sobre el sistema nervioso. Los estudios sobre la dinámica circulatoria, en los que los fisiólogos inyectaban sustancias inertes marcadas, permitieron estimar que la concentración inicial de la heroína intracerebral era enorme, peligrosamente próxima a un nivel letal. Si el sujeto sobrevive a este impacto se debe a la tolerancia alcanzada gracias a las dosis anteriores y a la brevedad de la punta narcótica. Por ejemplo, si un adicto se inyecta 20 miligramos, al cabo de unos diez segundos la concentración máxima circulante en el cerebro y otros órganos puede llegar a 40 microgramos por milimetro cúbico de sangre. Esta heroína permanecerá en contacto con células susceptibles durante un período de uno a tres segundos. En este breve tiempo, la droga ha sido arrastrada con la sangre venosa, a través del corazón derecho, a los pulmones, vuelve de nuevo al corazón y desde allí es bombeada a través de la sangre arterial hasta los tejidos sensibles. En este primer recorrido, la droga apenas se diluye en los líquidos corporales y el hígado no tiene la oportunidad de actuar sobre ella. A medida que el narcótico inicia su segundo recorrido orgánico, su concentración disminuye bruscamente. Cinco minutos después de la inyección puede haber descendido por debajo de los 0,5 microgramos por mililitro y, a las dos horas, es inferior a los 0,05 microgramos.

El adicto ha propinado, pues, a su sistema nervioso un verdadero martillazo farmacológico, que va seguido de una rápida abstinencia que sensibiliza al cerebro para la dosis siguiente. Estos efectos transitorios parecen constituir una parte importante del proceso adictivo: todos los adictos los buscan y no sólo los que se inyectan narcóticos, sino también las personas adictas a los sedantes (que se los inyectan cuando pueden), los alcohólicos (que echan un trago) y los fumadores (que inhalan profundamente el humo hasta los pulmones). Los cambios rápidos en la concentración de la droga potencian al máximo los efectos farmacológicos.

Cuando nuestro trabajo inicial con métodos analíticos no nos permitía controlar las concentraciones sanguíneas de narcóticos, Kreek y yo ya

observábamos los efectos funcionales agudos de narcóticos inyectados cuidadosamente en las venas de los adictos. A los diez segundos de iniciar la inyección de heroína, los sujetos presentaban la erupción narcótica típica, consistente en una oleada de sentimientos eufóricos, sensaciones viscerales, enrojecimiento facial y un timbre de voz más grave. Tras este efecto inmediato, que duraba sólo unos minutos, los sujetos volvían a un estado de calma, bienestar y despreocupación. Parecían completamente desinteresados por los acontecimientos del entorno durante la hora siguiente, aunque eran capaces de responder a las preguntas que se les formulaban.

Entre las dos y las cuatro horas posteriores a la inyección, reaparecía la inquietud. Se sentían incómodos y ansiaban otra inyección. La heroína los inutilizaba, primero, por sus efectos narcóticos y, después, por el síndrome de abstinencia. Ningún testigo de estos experimentos hubiese recomendado mantener a los adictos con heroína o con cualquier otra droga narcótica autoadministrable y de acción rápida. Los adictos que utilizaran la droga de esta manera no hubiesen podido vivir y trabajar como personas normales, ni tampoco mantener de forma constante una determinada dosis diaria.

El nivel sanguíneo de metadona, en cambio, nunca excede del microgramo por mililitro, ni desciende por debajo de 0,1, cuando se administran dosis orales a pacientes estabilizados. En la mayoría de ellos, la oscilación es incluso menor si la dosis se ajusta minuciosamente. Por término medio, la concentración se mantiene entre un máximo de 0,5 microgramos por mililitro y un mínimo de 0,2. La concentración máxima está muy por debajo del nivel necesario para provocar efectos narcóticos en los sujetos con tolerancia y, la mínima, supera sin peligro el umbral de la abstinencia. Además, las oscilaciones en la concentración son extraordinariamente lentas: el pico no se alcanza hasta 2 a 4 horas después de la ingesta de la dosis.

Tras la absorción de la metadona, un 98 por ciento desaparece de la sangre y se mantiene inactiva, de manera que sólo circula y alcanza a las células sensibles al narcótico un 2 por ciento. A medida que se excreta esta parte activa, la droga fijada a los tejidos se disocia de ellos, vuelve al torrente sanguíneo y mantiene sus niveles circulantes. Así, la acción sobre el tejido nervioso no es brusca. Por otra parte, el intestino y el hígado garantizan la ausencia de oscilaciones radicales en la concentración del narcótico.

Conviene señalar también que la ingesta oral permite una entrada más lenta del tóxico puesto que, antes de que alcance la circulación general, debe atravesar la pared intestinal y salvar el hígado, donde se opera una eliminación de casi el 85 por ciento. La droga se libera entonces lentamente a medida que la sangre circula una y otra vez a través del hígado. Por tanto, aunque la dosis de metadona se ingiriera de golpe, tardaría un par o más de horas en verterse al torrente sanguíneo.

No se sabe de ningún otro narcótico que combine tan eficazmente el tiempo de acción y la seguridad de cara a una terapia de mantenimiento. La morfina y drogas afines se absorben muy mal por vía intestinal, y no se unen lo bastante a los tejidos como para mantener una concentración sanguínea estable durante 24 horas. El acetilmetadol, fármaco químicamente afín a la metadona, presenta un período de acción más largo que el de ésta, pero da lugar a metabolitos altamente tóxicos. En cambio, los metabolitos de la metadona son farmacológicamente inactivos, de modo que nuestra única preocupación consiste en proporcionar una dosis diaria que reemplace a la cantidad total metabolizada.

Esta comprensión teórica es bastante reciente, y han sido muchos los pacientes tratados incorrectamente con metadona durante los últimos quince años. Algunos programas han empleado dosis inadecuadas (30 miligramos diarios o menos). Los pacientes atendidos según este protocolo debieron presentar seguramente síntomas recurrentes de abstinencia cuando la concentración de droga circulante descendía demasiado. Lo habitual en esta situación es que el adicto tenga la tentación de proporcionarse un suplemento diario con drogas ilegales. Algunos programas proporcionaban ampollas de metadona inyectable. La consecuencia de eliminar la barrera intestinal y hepática se resume en que el cerebro vuelve a recibir un martillazo similar al de la inyección de heroína.

El tratamiento con metadona tiene sus propias limitaciones, como se desprende del estudio de seguimiento llevado a cabo hace poco por Herman Joseph y el autor. Entre las 70.000 personas que habían iniciado programas de este género en Nueva York entre 1965 y 1976, seleccionamos una muestra al azar (estructurada apropiadamente para reflejar el sexo y la edad de los sujetos que participaban en tales programas) de 1513 pacientes, que representaban a los que habían empezado a comienzos del programa, en 1966-67, y a los incorporados al final, en 1972. En el grupo elegido, 667 estaban todavía en tratamiento y 864, no. Durante los tres años que siguieron, un grupo de profesionales expertos examinaron los registros de las agencias y entrevistaron por distintos motivos a 1427 individuos, es decir, el 94 por ciento de la muestra.

La conclusión más notable que se extrajo del conjunto de los datos fue el fracaso de la rehabilitación social para curar la adicción a narcóticos. Aunque el mantenimiento había reducido de manera notable el consumo de narcóticos ilegales y la delictividad asociada a ellos, y aunque la mayoría de los pacientes habían conservado su empleo mientras estaban en tratamiento, por lo general volvían a consumir narcóticos ilegales cuando cesaba el mantenimiento con metadona. En cambio, el grupo que estaba todavía en tratamiento no mostraba tendencia a la recaída.

No es esto lo que se esperaba cuando se empezó la investigación con metadona. Sea como fuere, los hechos no pueden ignorarse. Ningún programa aplicado a personas con un historial de profunda adicción a los narcóticos ha tenido más éxito que el de mantenimiento con metadona de Nueva York. Ningún programa (de mantenimiento o de supresión) que tratase a adictos de la misma gravedad con seguimientos de tres años ha presentado pruebas de mejores resultados a largo plazo. Si bien una minoría de sujetos con historiales de adicción grave puede mantenerse abstinente tras el alta (o sin tratamiento alguno), lo cierto es que la mayoría recae y la mayoría requiere un prolongado programa de mantenimiento para comportarse normalmente.

Nuestra experiencia de 15 años en el tratamiento de adictos nos proporciona datos para someter a prueba alguna de las antiguas teorías sobre la conducta adictiva. Los exámenes psiquiátricos de pacientes estabilizados en los programas de mantenimiento no han detectado ningún rasgo común que permita hablar de una personalidad adictiva. Al igual que los adictos que emplean drogas ilegales, las personas que acudieron a los programas de mantenimiento habían tenido experiencias similares y, en algunos aspectos, incluso actitudes parejas; pero parece claro que las experiencias y las actitudes eran más consecuencia que causa de la adicción. Con tratamiento médico y rehabilitación social, estas personas se

volvían tan diferentes, desde un enfoque psicológico, como sus conciudadanos no adictos.

También pudimos observar que en pacientes farmacológicamente estabilizados no se detectaban las anómalas búsquedas de placer o las evasiones de la realidad que se han atribuido a los adictos. Cualquiera que fuera el grado de euforia con que el adicto respondiera a la heroína antes del tratamiento, desaparecía por completo al estabilizarse con metadona, e incluso en la mayoría de los casos el tratamiento era bien recibido. Los adictos con historiales de dos o más años de sometimiento a la heroína parecían completamente satisfechos de sacrificar esa euforia ocasional en favor de una continua sensación de normalidad. Este hecho sugiere que la conducta adictiva del consumidor inveterado está más motivada por aliviar el malestar que por perseguir el placer. Lo mismo puede aplicarse a las personas incapaces de prescindir de otros hábitos nocivos, como fumar o beber en exceso.

En tanto en cuanto la adicción es un síntoma de disforia, la práctica nociva debería suplantarse por una medicina alternativa de supresión de síntomas, siempre, claro está, que la alternativa fuera segura y aceptable. Aunque tal estrategia no implica una cura absoluta, puede salvar miles de vidas hasta que los investigadores posean un conocimiento teórico completo de la conducta adictiva. Además, la notable diferencia entre los efectos del consumo continuo de metadona y de heroína proporcionan una pista que considero importante respecto al problema fundamental: el sistema nervioso responde sobre todo a los cambios de concentración de los compuestos circulantes neuroactivos y tiende a volverse insensible cuando las concentraciones se mantienen uniformes. Los adictos buscan la droga para conseguir una concentración cerebral bruscamente elevada, puesto que así se obtiene la subida (también llamada "high" o "flash"). La introducción de la ginebra en el siglo XVIII fue seguida de una epidemia de alcoholismo, la introducción de la aguja hipodérmica en el siglo XIX, de una epidemia de adicción a los narcóticos y la introducción del cigarrillo en este siglo, de una epidemia de fumadores. En términos técnicos, la investigación futura deberá tener en cuenta la farmacocinética de las drogas adictivas –la variable tiempo– puesto que la sensibilidad del sistema nervioso para responder cambia según el tiempo de contacto con el tóxico.

El sistema de huida de la cucaracha

¿Cómo evade este insecto el golpe de una persona o la acometida de un sapo? Los mecanismos de los que se vale constituyen un modelo simplificado de los que rigen en formas de conducta desarrolladas por especies más complejas

Jeffrey M. Camhi

La dueña de la casa observa que la cucaracha del estante de la cocina es una experta en librarse de un zapatillazo mortal. El naturalista observa que la cucaracha del suelo del bosque escapa las más de las veces al latigazo de la pegajosa lengua de un sapo. ¿De qué modo el sistema nervioso del insecto inicia y controla dichos actos de comportamiento? En última instancia, una pregunta de este tipo puede llevarle a uno a especular acerca de la base neuronal del comportamiento de animales más complejos, sin excluir al hombre. Pero intimidados por la complejidad del cerebro humano, muchos investigadores se han sentido atraídos por sujetos experimentales mucho más simples: crustáceos, moluscos e insectos. Estos animales tienen relativamente pocas neuronas, o células nerviosas, y muchas de ellas son grandes y reconocibles una a una. Ello nos permite determinar, a veces, el papel que cada neurona desempeña en un acto de comportamiento, importante información que es casi imposible de obtener en estudios del cerebro de mamíferos.

Veamos un ejemplo de las técnicas empleadas en tales experimentos: el investigador pincha una neurona con un electrodo introducido en una micropipeta de cristal y registra las señales de voltaje que se generan en la membrana de la célula. Un tipo de señal es el potencial de acción, diferencia de potencial o voltaje que se propaga a lo largo de un axón (la fibra larga de la neurona) y sirve para la comunicación a larga distancia en el sistema nervioso. Pueden también registrarse potenciales sinápticos, las diferencias de potencial que se generan localmente en una sinapsis, o punto de comunicación funcional entre dos neuronas.

Para saber si dos neuronas cualesquiera se comunican o no entre sí, el investigador inserta microelectrodos en ambas células y determina si al potencial de acción de una célula sigue invariablemente un potencial sináptico en la otra neurona. El potencial sináptico puede ser excitador, es decir, que aumenta la probabilidad de que la segunda célula genere a su vez un potencial de acción, o inhibidor, o sea, que reduce momentáneamente la probabilidad de que la segunda célula se dispare. Repitiendo el experimento en distintos pares de neuronas que originen un acto de comportamiento, se obtiene en último término la información necesaria para trazar un diagrama de circuito: la pauta de interacciones sinápticas entre las neuronas identificadas que intervienen en el comportamiento. Un diagrama de este tipo se parece mucho al esquema que esboza un ingeniero de los componentes electrónicos y de sus conexiones de una radio o de una calculadora electrónica. Se han establecido ya los diagramas de circuito de algunas pautas de comportamiento de animales invertebrados.

Buena parte de este tipo de trabajo ha puesto más énfasis en la neurología subyacente al comportamiento que en el propio comportamiento. Ello es lamentable, porque, para determinados tipos de comportamiento, la consideración atenta de los actos del animal conforme éstos se van desarrollando en condiciones naturales puede iluminar importantes características del control del comportamiento. El estudio conjunto del comportamiento de un animal en condiciones naturales y del control neuronal de dicha conducta ha venido en llamarse neuroetología.

Durante los últimos seis años, mis alumnos, investigadores de postdoctorado y yo mismo, en la Universidad de Cornell, hemos abordado, desde un enfoque neuroetológico, el estudio de un sujeto particularmente idóneo, la cucaracha americana, *Periplaneta americana.* Aunque se trata de un insecto originario del Africa tropical, en los últimos siglos su hábitat se ha extendido hasta incluir la mayor parte de las regiones tropicales, subtropicales y templadas del mundo. Este animal es familiar, incluso demasiado, en su papel de coinquilino de los hogares humanos. Hemos venido examinando el sistema de receptores sensoriales y las redes de integración de señales que permiten a la cucaracha escapar a las acometidas de sus depredadores naturales.

De entre quienes nos precedieron en la investigación, merece destacarse Kenneth D. Roeder, de la Universidad de Tufts. A finales de la década de 1940 señaló que las cucarachas corrían en respuesta a un ligero soplo de aire y que la respuesta desaparecía si se extirpaban los dos apéndices sensoriales posteriores del insecto, los cercos. Sobre la cara inferior de cada cerco Roeder observó una serie de pelos delicados, unos 220 de ellos en el insecto adulto. Más tarde se demostró que es su estimulación por el viento lo que hace que la cucaracha corra.

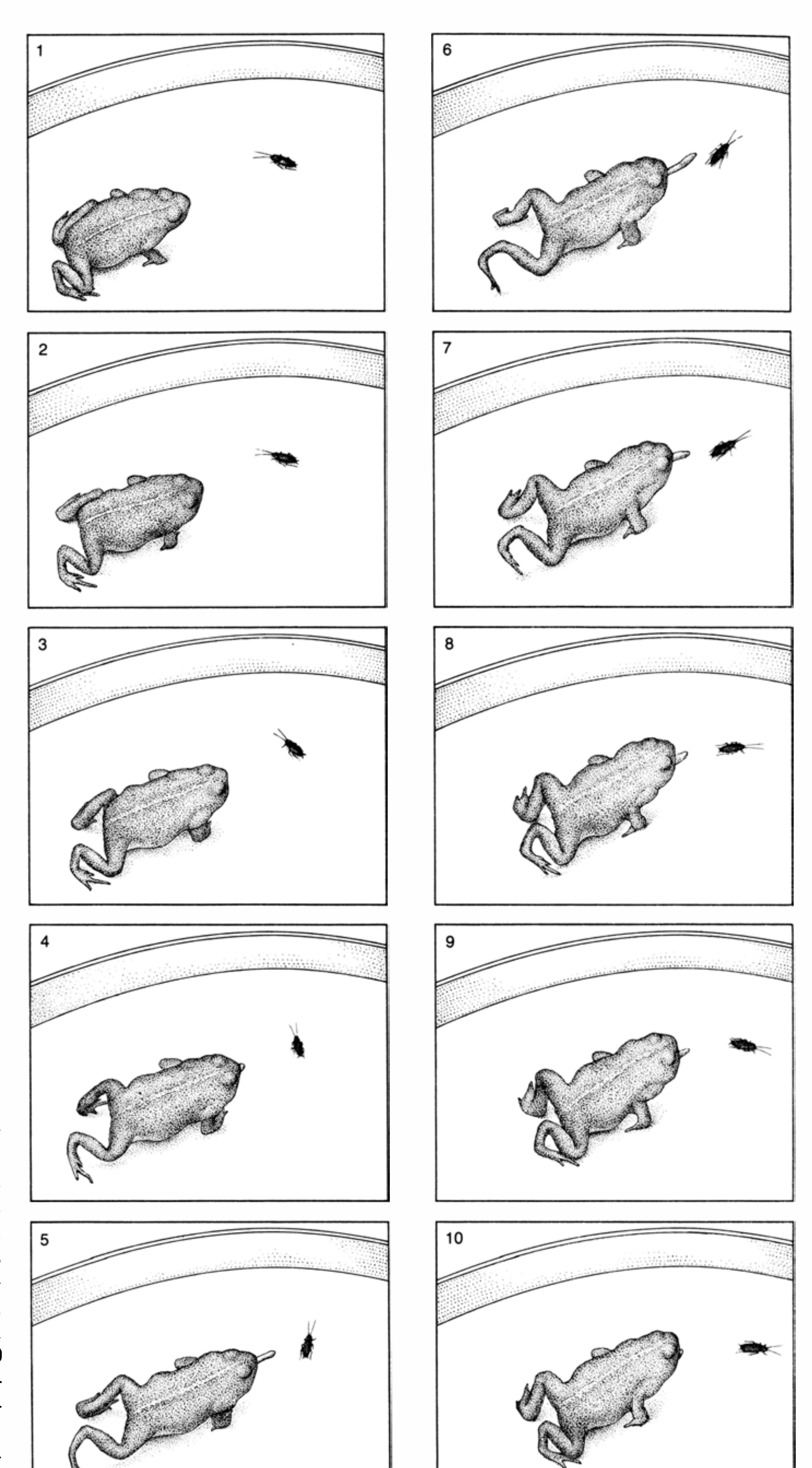
Cuando la cucaracha recibe un soplo de aire, muchos de los pelos cercales se doblan en su articulación. La deflexión de un pelo excita la neurona que se encuentra en su base. Cada neurona sensitiva se halla conectada al sistema nervioso central mediante su axón. Roeder demostró que, al excitarse las células sensoriales de los cercos, éstas excitaban a su vez a un grupo de células nerviosas particularmente grandes, de-

nominadas interneuronas gigantes. Una interneurona es una célula nerviosa que sólo se encuentra en el sistema nervioso central, donde actúa de eslabón funcional entre las neuronas sensitivas y las motoras. La cucaracha *P. americana* tiene 14 interneuronas gigantes, que podemos identificar a través del microscopio, en todos los miembros de la especie.

El cuerpo celular de cada interneurona gigante se halla en el ganglio terminal. En el sistema nervioso central de los invertebrados, un ganglio está compuesto por un grupo de cuerpos celulares y su entramado de ramas cortas, las dendritas, junto con axones cuyos terminales forman sinapsis con las dendritas. Entre ganglios cercanos corren estructuras conectivas similares a cables que sólo contienen axones. Los ganglios y las estructuras conectivas constituyen conjuntamente el cordón nervioso central del insecto. Tal como demostraron Micha Spira e Itzhac Parnas, de la Universidad Hebrea de Jerusalén, los axones de las interneuronas gigantes recorren todo el cordón nervioso, hasta llegar a la cabeza. Los potenciales de acción ascienden por los axones de las interneuronas gigantes y pasan a través de los ganglios torácicos, donde se encuentran las neuronas motoras que controlan las patas.

Roeder súgirió que las interneuronas gigantes excitaban a las neuronas motoras de la pata de la cucaracha, ya directamente, ya a través de otras interneuronas. Roy E. Ritzmann, que trabajó en mi laboratorio y se halla ahora en la Universidad Case Western Reserve, demostró que cuando se estimulaba una interneurona gigante mediante un microelectrodo intracelular para que generase una secuencia de potenciales de acción, se activaban las neuronas motoras de la pata. Estos experimentos, y otros afines, nos han convencido de que las interneuronas gigantes facilitan la carrera en respuesta a la estimulación de los cercos por el viento. Resumiendo, sabemos que el circuito de huida consta de unos 440 receptores de viento cercales, 14 interneuronas gigantes y un reducido número de neuronas motoras de las patas.

Pretendíamos descubrir de qué modo estas células y sus interconexiones producían el comportamiento de carrera evocado por el viento. Sin embargo, al principio, no se sabía casi nada del comportamiento, salvo su existencia. Roeder había propuesto que esa estrategia podría ayudar a la cucaracha a escapar de aquellos depredadores cuya



ESCAPAR POR UN PELO. Una cucaracha (*Periplaneta americana*) huye del ataque de un sapo (*Bufo marinus*), según se indica en estas escenas redibujadas a partir de fotogramas de una película de alta velocidad realizada en el laboratorio del autor, en la Universidad de Cornell. Después de efectuar una corta acometida, el sapo intenta atrapar al insecto con su lengua pegajosa. El movimiento del cuerpo del sapo crea un soplo de aire que la cucaracha percibe mediante los pelos que sobresalen de dos apéndices sensoriales posteriores, denominados cercos. La cucaracha responde dando la vuelta y apartándose a la carrera del origen del viento. En condiciones óptimas, la respuesta de la cucaracha puede iniciarse a los 11 milisegundos de haberse activado los pelos cercales. La secuencia se filmó a 64 imágenes por segundo.

acometida pusiera en movimiento el aire circundante. Razonamos que si ésta era la función real del comportamiento, debería ser direccional, habida cuenta de que el insecto tendría su mayor probabilidad de supervivencia en alejarse del origen del viento.

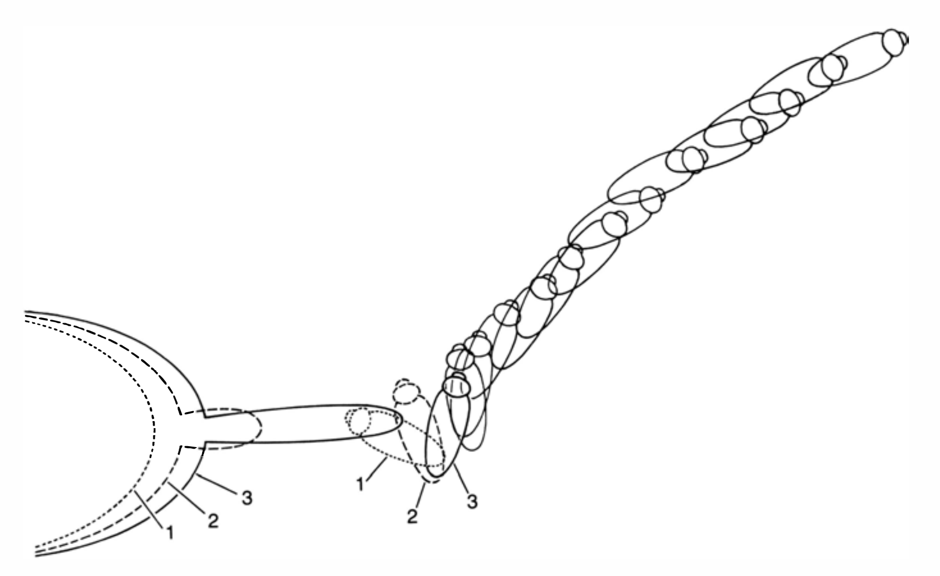
Para comprobar la direccionalidad, Winston Tom y el autor filmaron películas de alta velocidad de cucarachas que recibían soplos de viento normalizados dirigidos desde distintos ángulos. Al representar gráficamente, tras muchos ensayos, la dirección en que los animales se daban vuelta y corrían, resultó claro que, por lo general, los insectos se enfilaban en el sentido contrario al del origen del viento. Además, no cabía duda de que los responsables de la respuesta orientada eran los pelos receptores de viento de la cara inferior de los cercos. Esta conclusión se basaba en varios tipos distintos de pruebas.

Cuando cubrimos la cara inferior de ambos cercos con cera, las cucarachas no respondieron en absoluto a los estímulos de viento. Si se cubría la cara superior de los cercos, la respuesta direccional no resultaba afectada. De girarse ambos cercos en el sentido de las agujas del reloj, a nivel de donde se hallan articulados al cuerpo, fijando cada cerco en su lugar con una pizca de cera, encontrábamos que la respuesta direccional al viento variaba en, aproximadamente, la cantidad correspondiente al giro. Cuando cubríamos un cerco o lo arrancábamos, la mayoría de las vueltas que el insecto efectuaba eran en la dirección opuesta a la del cerco intacto, con independencia de la dirección del viento.

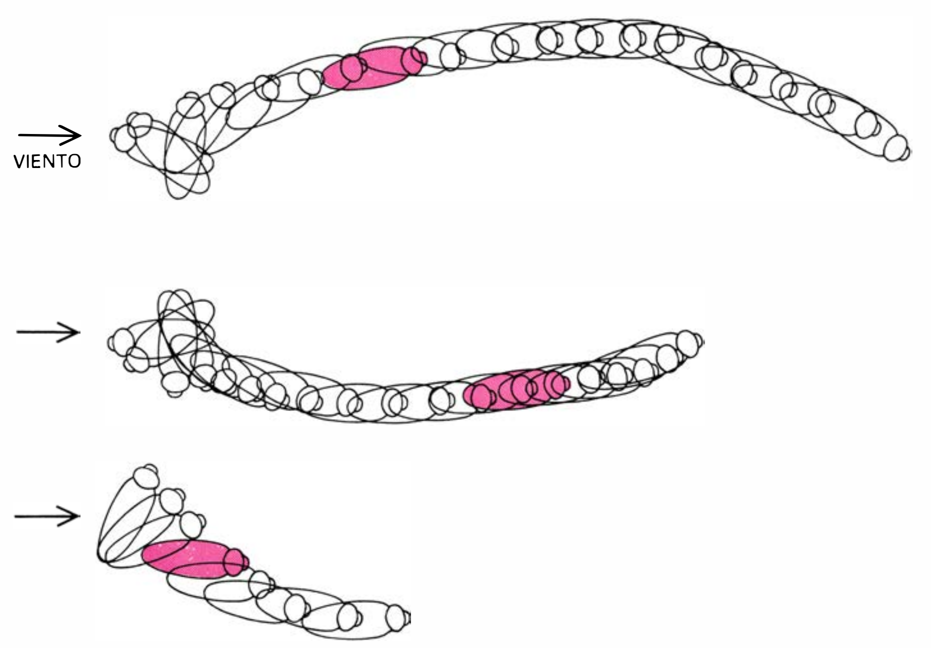
Sometimos a prueba la respuesta de huida en cucarachas juveniles del primer estadio de muda. Aunque los juveniles sólo tienen dos pelos en cada cerco, se apartan de una fuente de viento casi con el mismo éxito que los adultos. Además, como ha demostrado Susan Volman, de nuestro laboratorio, si se arrancan los dos pelos de un cerco los insectos juveniles se apartan del lado que tiene el cerco intacto, independientemente de la dirección del viento. Si se eliminan los pelos de ambos cercos, los insectos ya no responden a los soplos de aire. Diversos experimentos de control nos convencieron de que el estímulo responsable de las respuestas orientadas era en verdad el viento y no las señales visuales, auditivas u otras que pudieran proceder del simulador de viento.

El paso siguiente fue convencernos de que las cucarachas se basaban realmente en el comportamiento de giro y carrera evocado por el viento para huir de los depredadores naturales. El depredador que elegimos para el estudio fue el sapo tropical *Bufo marinus,* que es activo por la noche y del que se sabe que en su hábitat natural acomete contra las cucarachas nocturnas. Empezamos por adaptar un sapo y una cucaracha a las condiciones de iluminación nocturna del laboratorio. Colocamos luego a ambos animales en una cámara de un metro de diámetro. En su desplazamiento por la misma, la cucaracha terminaba por acercarse lo suficiente (alrededor de los 15 centímetros) al sapo como para que éste emprendiese una acometida depredadora, es decir, un salto hacia adelante que culminaba en un latigazo de su pegajosa lengua.

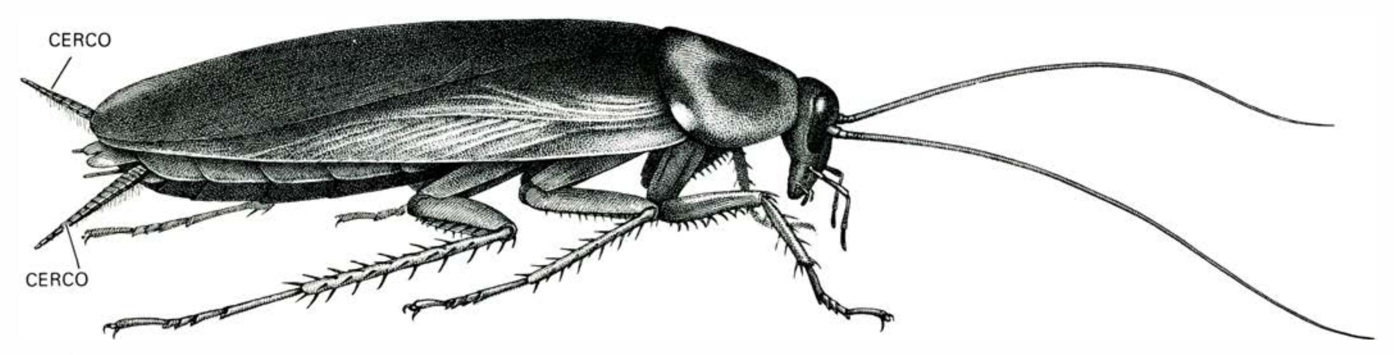
Observando muchas acometidas realizadas por varios sapos a muchas cucarachas distintas, determinamos que los insectos escapaban con éxito en un 55 por ciento de los ataques. En cambio, las cucarachas a las que se



HUIDA DE UNA CUCARACHA del ataque de un sapo, en detalle. Los dibujos se fundan en 15 fotogramas sucesivos de película. La secuencia comienza con la cabeza del sapo y el insecto en la posición *1*. En el segundo fotograma de la película (*2*), el sapo ha empezado a extender la lengua. En el tercer fotograma (*3*) la lengua del sapo se halla completamente extendida y yerra por poco al insecto, que se da la vuelta y continúa corriendo. La posición de la cucaracha se indica a intervalos de 16 milisegundos.



RESPUESTAS DIRECCIONALES del sistema de huida de la cucaracha según puede verse en tres ejemplos de un ensayo en el que se dirigieron soplos controlados de aire contra el insecto desde direcciones distintas. La silueta situada más a la izquierda, la más cercana a la flecha, muestra la posición de la cucaracha justo antes del estímulo. Las siluetas sucesivas muestran la posición del insecto a intervalos de 16 milisegundos. La silueta coloreada de cada caso indica la posición en el momento de completar su giro inicial. Los fotogramas sucesivos no manifiestan ningún giro adicional en la dirección de partida.



PERIPLANETA AMERICANA, una de las cucarachas más comunes en los hogares americanos y europeos, es nativa de Africa tropical. La figura ilustra un macho adulto, cuyo cuerpo tiene unos 40 milímetros de longitud. Los receptores sensibles al viento son unos 440 pelos, situados en los cercos.

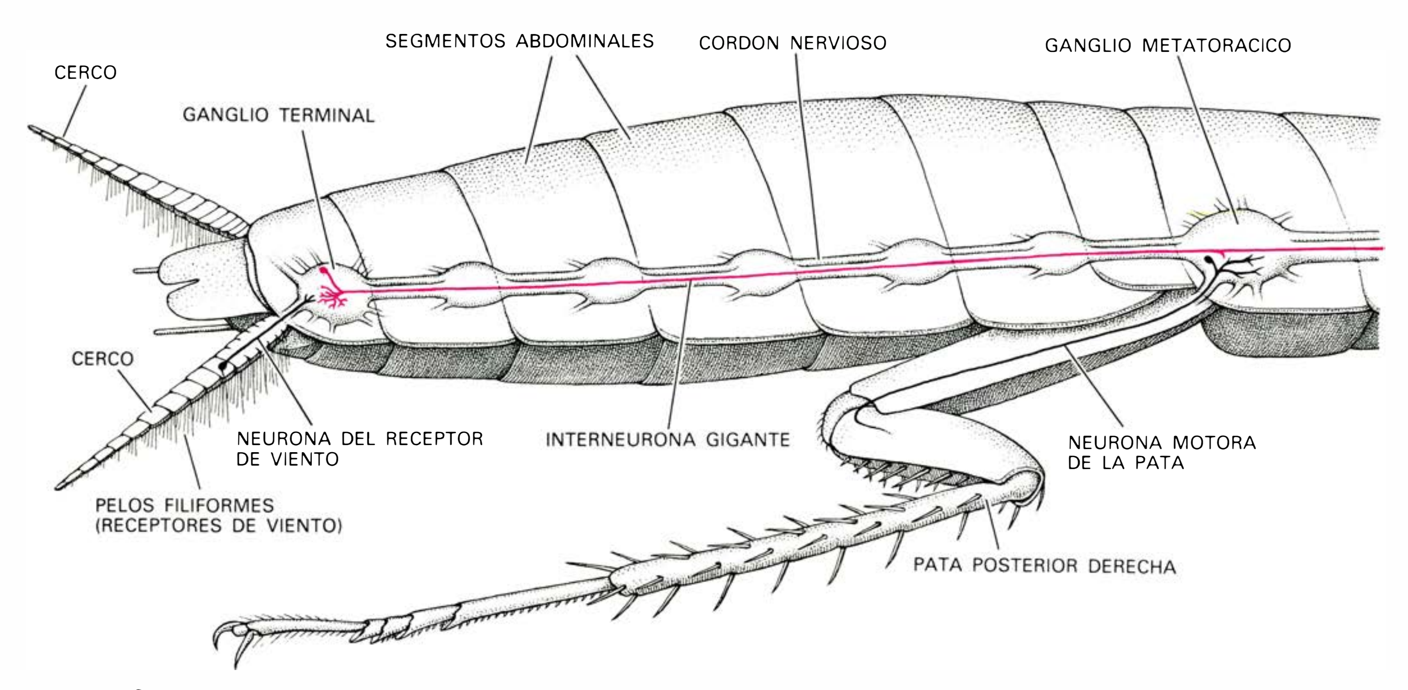
había cubierto los cercos con cera escaparon sólo el 8 por ciento de las veces. Los insectos de un grupo de control, a los que habíamos cubierto parte del abdomen pero no los cercos, escaparon del 47 por ciento de los ataques. No había duda de que los cercos ayudaban a las cucarachas a escapar de los sapos, seguramente mediante algún tipo de receptor cercal. Aunque los receptores de viento eran los candidatos más plausibles, en ese estadio no podíamos excluir la posibilidad de que hubiera receptores cercales capaces de detectar las vibraciones del suelo, el sonido, los olores u otras señales procedentes del sapo.

A tal fin sacamos películas de alta velocidad de los encuentros entre las cucarachas y los sapos. Las películas demostraron de manera inequívoca que los insectos respondían a los sapos con las mismas pautas que a un soplo de viento: dándose la vuelta y corriendo. Nos convencimos sin la menor reserva de que el viento era la señal crucial al medir el viento creado por la acometida del sapo y compararlo con la cantidad de viento necesaria para hacer que el insecto corriera.

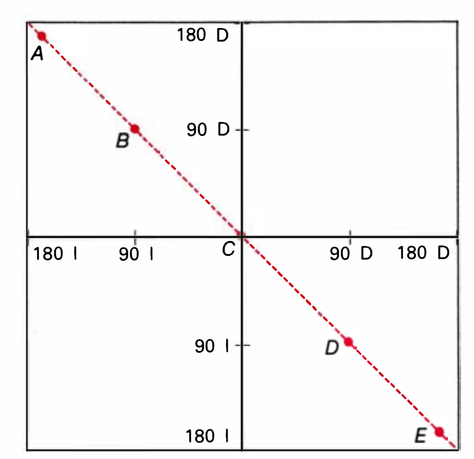
Para medir el viento generado por un sapo que acomete, anestesiamos a una cucaracha con dióxido de carbono, la sujetamos al extremo de una pequeña caña de pescar y la soltamos sobre el suelo de la cámara que contenía un sapo. (Los sapos acometen únicamente a los objetos que se mueven o que acaban de moverse.) Tan pronto habíamos captado la atención del sapo, situábamos la cucaracha a un centímetro del elemento activo de un anemómetro que había sido colocado sobre el suelo, en el centro de la cámara. Cuando el sapo acometía se registraba la intensidad del viento producido por la embestida.

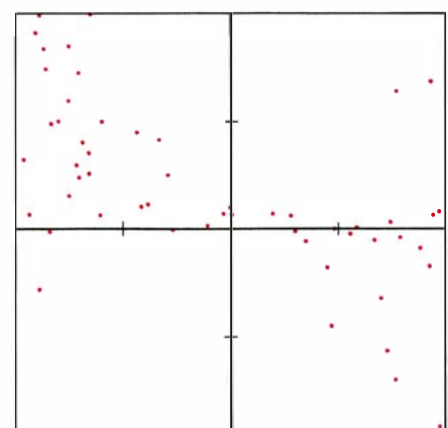
Puesto que, simultáneamente, filmábamos la acometida, pudimos relacionar la velocidad del viento con la distancia entre el sapo y la cucaracha en cada momento. A partir de películas anteriores sobre sapos que atacaban a cucarachas sin anestesiar, habíamos deducido que los insectos iniciaban el movimiento de huida, por término medio, alrededor de un fotograma (o unos 16 milisegundos a una velocidad de filmación de 64 fotogramas por segundo) antes de que la lengua del sapo saliera de su boca. La intensidad del viento justo antes de este momento era, por tanto, de un interés crucial.

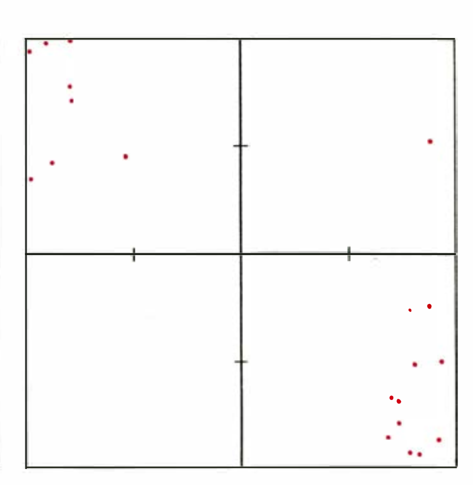
En otra serie de ensayos con soplos controlados de viento procedentes de nuestro simulador, Mark Plummer y el autor encontraron que, para provocar la carrera de los insectos, tan importante era una elevada aceleración del viento como una determinada velocidad crítica. En concreto, para evocar

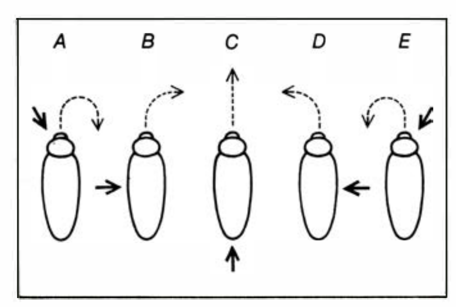


LAS NEURONAS DEL SISTEMA DE HUIDA están organizadas de modo que el estímulo que produce un soplo de aire provoque el giro y la carrera. Los impulsos nerviosos que se originan en las neuronas del receptor de viento son reenviados desde el ganglio terminal hasta las interneuronas gigantes que discurren por el cordón nervioso central, y de allí al ganglio metatorácico, donde se integra la señal que excitará las neuronas motoras de las patas.









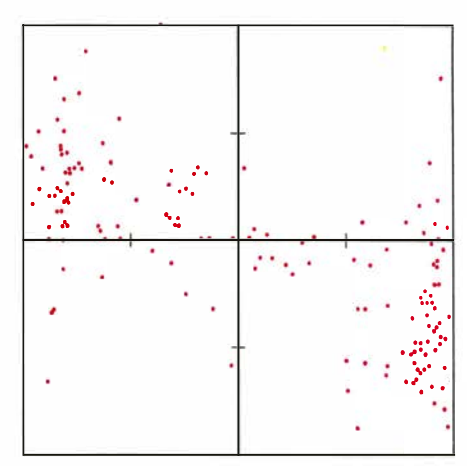
RESPUESTAS DE GIRO DE LAS CUCARACHAS representadas sobre gráficos (tal como se realizó en el laboratorio del autor), según las convenciones que se ilustran a la izquierda. Las cinco siluetas de insectos muestran la dirección del viento (*flechas continuas*) y la dirección del giro del insecto (*flechas a trazos*). Un ensayo experimental, donde el viento procede de atrás y la cucaracha responde corriendo hacia adelante, se representaría mediante el punto *C* del gráfico (tanto el estímulo como la respuesta se hallan a 0 grados). El viento procedente de 90 grados a la izquierda que evocara una vuelta de 90 grados a la derecha se representaría mediante el punto *B*, y así sucesivamente. En el gráfico de la izquierda, la línea inclinada de color representa la localización de los puntos resultante si las cucarachas se orientaran perfectamente en sentido contrario al del estímulo del viento. El gráfico central muestra las respuestas de giro de cucarachas macho adultas a soplos de aire del laboratorio. Su tendencia a dirigirse en una dirección contraria a la del viento es altamente significativa. El gráfico de la derecha representa las respuestas de giro de cucarachas macho adultas ante las acometidas de sapos, dibujadas a partir de secuencias filmadas. Sólo se registran los giros de los insectos que consiguieron sobrevivir a los ataques.

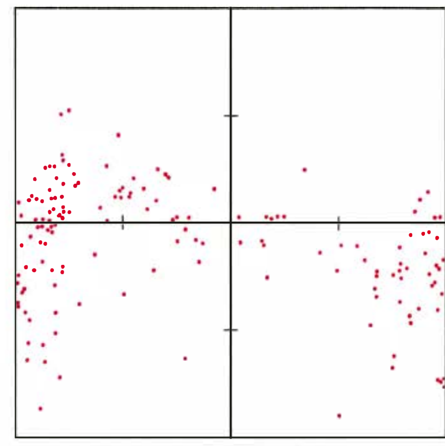
una respuesta de carrera en más de la mitad de los ensayos el viento había de mantener una aceleración de por lo menos 60 centímetros por segundo en cada segundo, combinada con una velocidad de por lo menos 12 milímetros por segundo. La latencia media de la respuesta, es decir, el tiempo medio que transcurría desde la llegada del soplo de aire a los cercos hasta el inicio de la carrera, era de 44 milisegundos.

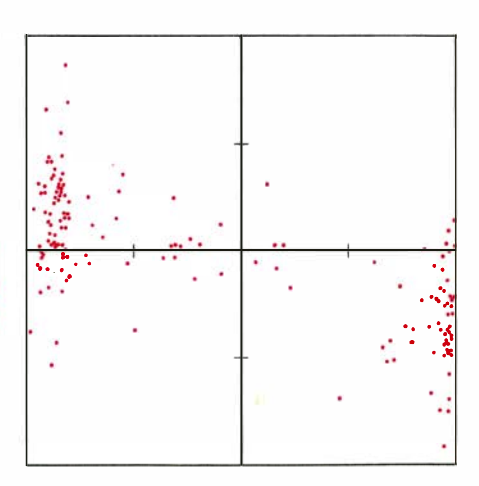
En los ataques del sapo a cucarachas anestesiadas, el viento producido por el sapo alcanzaba la aceleración y la velocidad críticas, por término medio, unos 41 milisegundos antes de que las cucarachas arrancaran en respuesta a la acometida del sapo. La concordancia entre los dos valores (44 milisegundos para la latencia y 41 milisegundos antes del movimiento) valía de prueba fehaciente para mostrar que el soplo de aire generado por la embestida del sapo era estímulo suficiente para disparar la respuesta de huida.

Fueron muchos los intentos realizados en busca de otras señales eficaces que el sapo pudiera producir; fracasaron todos. Por ejemplo, si disponíamos dos delgadas láminas de un material acrílico transparente entre el sapo y varias cucarachas en movimiento, los insectos permanecían indiferentes ante la visión, el sonido y la vibración que hacía el anfibio cuando, al acometer, chocaba con la barrera situada a menos de un centímetro de distancia de éstos. Otros experimentos nos permitieron descartar que las señales olfativas y táctiles contribuían de manera significativa a la respuesta de huida.

Por todo ello, podemos decir con cierta seguridad que el soplo de aire, breve pero intenso, que provoca la acometida del sapo es el principal canal de información, si no el único, a través del cual la cucaracha detecta a un sapo que ataca. Aunque suele considerarse







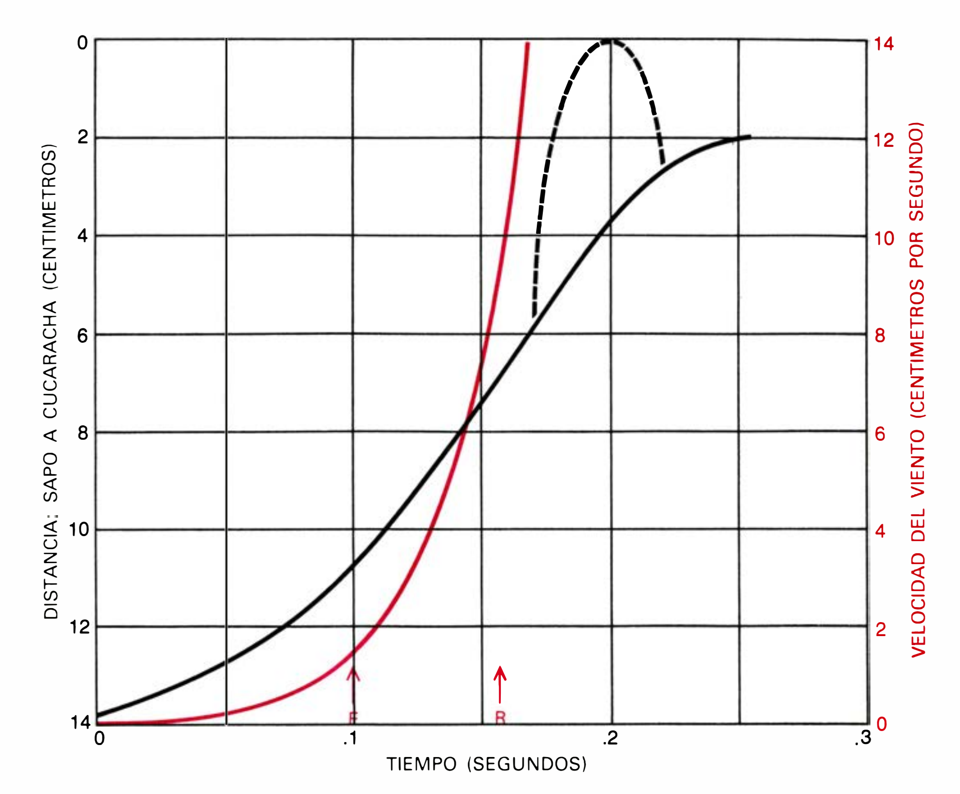
RECUPERACION de la extirpación de los cercos, según puede observarse en este conjunto de tres gráficos. El de la izquierda muestra las respuestas a soplos de aire de cucarachas ninfales normales, en un estadio de desarrollo avanzado. El gráfico del centro refleja las respuestas de insectos ninfales a los que se había extirpado el cerco izquierdo un día antes de ser sometidas a prueba. Las respuestas de giro a soplos de aire procedentes de la izquierda se distribuyeron al azar. La evitación de los soplos de aire procedentes de la derecha se mantuvo aproximadamente normal. El gráfico de la derecha muestra los resultados cuando se ensayó con los mismos insectos 30 días más tarde. Habían recuperado ya gran parte de la capacidad de ponerse en fuga.

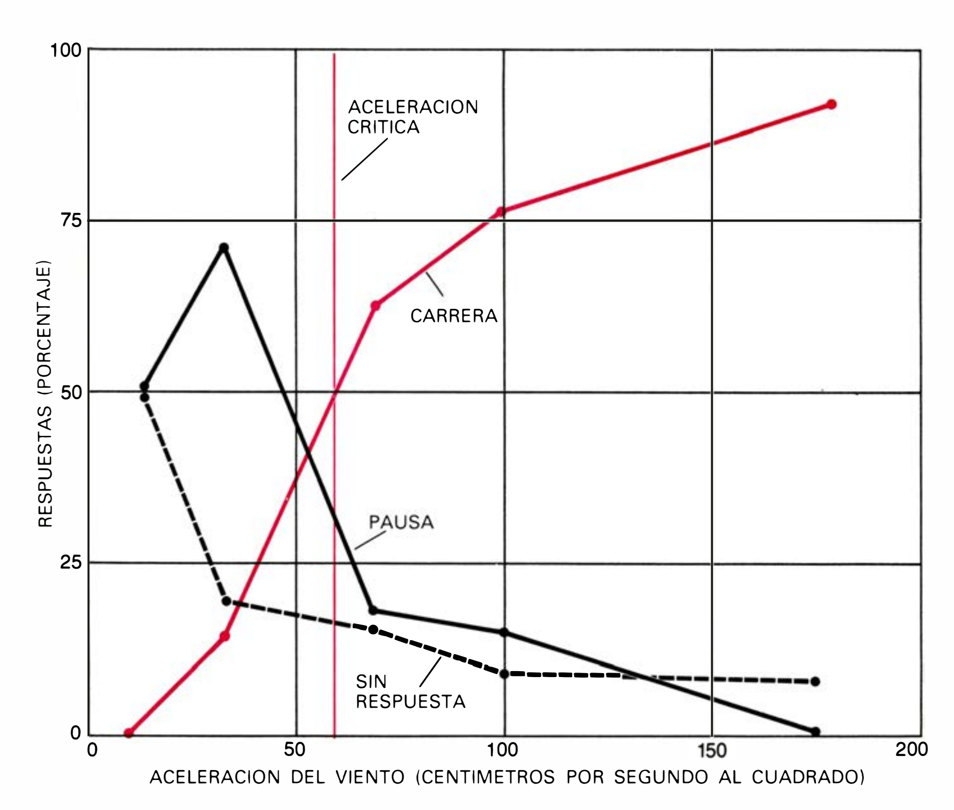
que el viento no contiene mucha información, en este caso sí la encierra y le anuncia al insecto la presencia y la dirección de un objeto que se acerca rápidamente con peligro para su vida.

La información relativa a la dirección de un soplo de aire debe estar codificada, de algún modo, en los receptores sensoriales de los cercos. El mecanismo básico lo descubrió, a finales de la década de 1960, R. Nicklaus, de la Universidad de Munich. Los pelos de un cerco se disponen en filas y columnas. Cada uno de los 17 segmentos cercales del adulto, excepto los situados en la punta o en torno a la base del cerco, tienen una fila de nueve pelos particularmente largos. Los pelos situados de forma similar en cada segmento pueden considerarse miembros de una columna que corre desde la parte anterior hasta la posterior del cerco. Todos los pelos de una columna determinada son iguales desde un punto de vista mecánico, porque pueden desviarse fácilmente en los dos sentidos de una dirección y menos fácilmente en ángulo recto a esta dirección. La dirección de máxima flexibilidad difiere de una columna a otra.

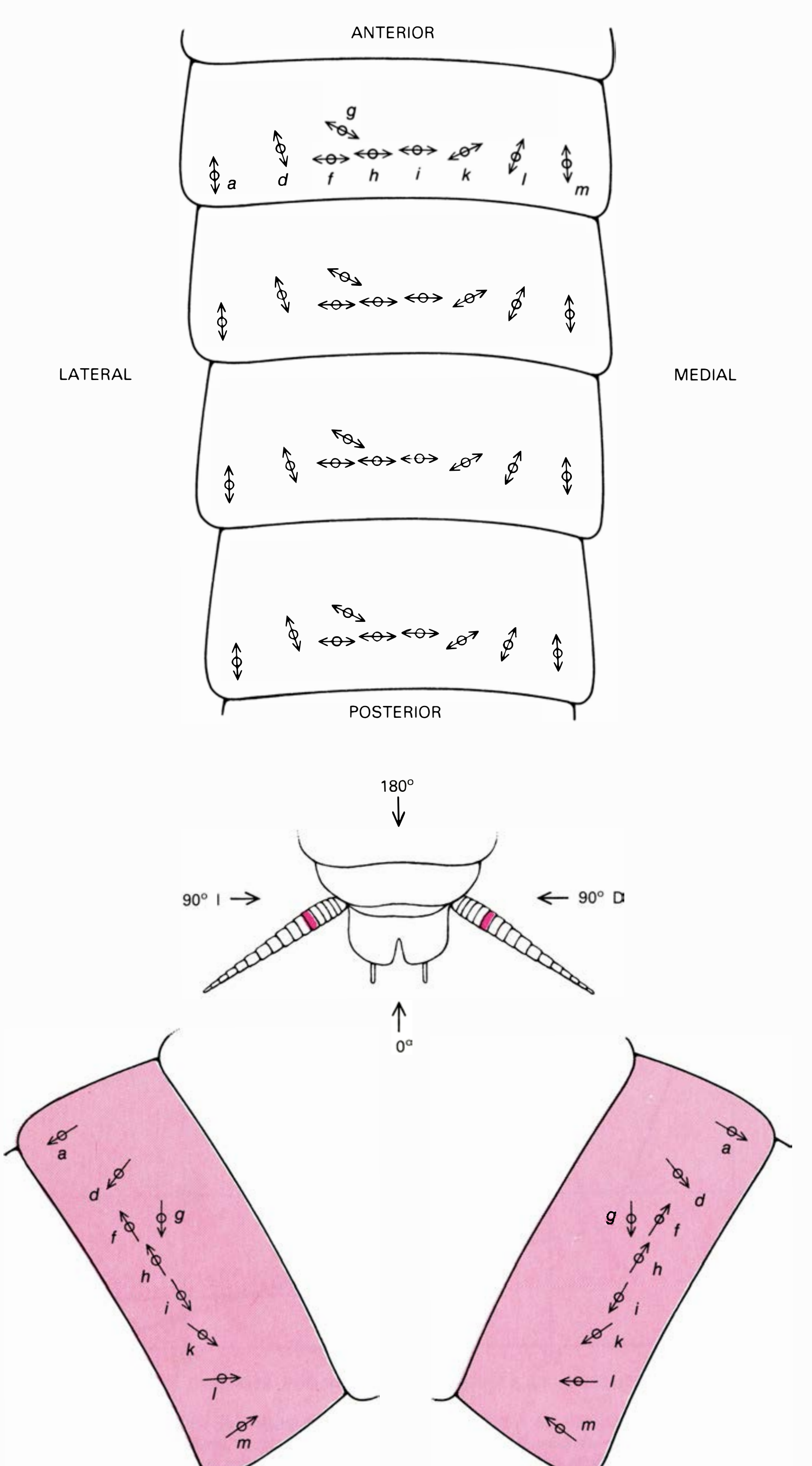
Registrando la salida sensorial de unos cuantos pelos individuales, Nicklaus observó que la desviación en un sentido de la dirección más flexible evocaba un número máximo de potenciales de acción. La inclinación en el sentido contrario (es decir, a 180 grados de distancia del primero) producía una inhibición máxima de los potenciales de acción de ocurrencia espontánea. Se supone que la inclinación excitadora estira la dendrita de la célula sensitiva, lo que provoca la generación de potenciales de acción. La desviación en sentido opuesto, inhibidora, parece relajar la dendrita de la célula sensitiva.

Daniel Dagan amplió, en mi laboratorio, estos hallazgos al determinar la dirección de excitación óptima de cada una de las nueve columnas de pelos cercales. Su técnica consistía en recubrir con cola todos los pelos, menos los de una columna, y registrar luego las respuestas de todo el nervio cercal a viento procedente de distintas direcciones. Observó que las nueve columnas cubrían, en su conjunto y de un modo eficiente, la rosa de los vientos. Mientras tanto, Joanne Westin, trabajando primero en mi laboratorio y posteriormente en Case Western Reserve, realizó registros intracelulares de axones aislados que transportaban impulsos sensoriales procedentes de pelos cercales, individualmente considerados.





VIENTO PRODUCIDO POR LA ACOMETIDA DE UN SAPO, registrado utilizando como blanco una cucaracha anestesiada y un anemómetro sensible. La curva en negro de arriba representa la distancia desde la parte anterior de la cabeza del sapo hasta la cucaracha durante el curso de la acometida. La lengua del sapo (*línea de trazos*) emerge hacia los 0,18 segundos de iniciarse la acometida. La curva en color representa la velocidad en el blanco del viento creado por la embestida del sapo. El viento alcanza una velocidad máxima de unos 20 centímetros por segundo (muy por encima de la escala) a unos 0,2 segundos del inicio de la acometida. La flecha *E* representa el momento en el que el viento alcanzó una aceleración crítica de 60 centímetros por segundo en cada segundo. La flecha *R* marca el tiempo medio (tomado en pruebas con insectos completamente alerta) en que las cucarachas comenzaron a moverse en respuesta a las acometidas del sapo. (En tales pruebas, *R* se mide con relación al momento de emergencia de la lengua del sapo.) En esta muestra del ensayo, el tiempo transcurrido entre *E* y *R* es de 56 milisegundos. Las tres curvas de abajo muestran respuestas de cucarachas en movimiento a soplos de aire de distintas aceleraciones, pero todas de la misma velocidad punta: cuatro centímetros por segundo. Según la aceleración, la respuesta es correr, realizar una pausa o bien ignorar el estímulo. Por encima de una aceleración de 60 centímetros por segundo en cada segundo más de la mitad de los insectos corren.



PELOS RECEPTORES DE VIENTO que se encuentran primariamente sobre la cara inferior de los dos cercos de una cucaracha. El dibujo de arriba identifica los pelos más largos sobre cuatro segmentos de un cerco de una cucaracha adulta. Cada segmento tiene un conjunto idéntico de nueve pelos (*a, d, f, g, h, i, k, l, m*), que se inclinan hacia delante y hacia atrás más fácilmente en una dirección que en cualquier otra. Las flechas de doble punta señalan la dirección de máxima flexibilidad hacia delante y hacia atrás para cada pelo. Uno de los dos sentidos de la dirección de menor resistencia corresponde asimismo al sentido de máxima respuesta excitadora a los soplos de aire. En la ilustración de abajo, las flechas de una sola punta indican el sentido de máxima respuesta al viento para cada pelo en dos segmentos cercales, uno del cerco izquierdo y otro del derecho. Hemos dibujado los cercos como si fueran transparentes, viéndose los pelos desde arriba. El viento que incide sobre un pelo desde un punto opuesto al sentido de la flecha produce una inhibición máxima de los potenciales de acción, que de otro modo surgirían espontáneamente.

Sus resultados mostraban que cada axón sensitivo respondía al máximo en un espectro bastante estrecho de direcciones del viento, y que, tomados en conjunto, los pelos ofrecían una cobertura de 360 grados.

¿Preservan la información direccional codificada por las células sensitivas las interneuronas gigantes excitadas por los receptores cercales? Westin, Jonathan Langberg y el autor respondieron a esta cuestión tomando registros intracelulares de interneuronas gigantes aisladas. Después de cada experimento, la interneurona gigante con que se había ensayado se identificaba bajo el microscopio de acuerdo con sus características anatómicas distintivas. La anatomía de la célula se hacía evidente inyectando un colorante a través del electrodo micropipeta que se había utilizado para el registro. El colorante inyectado (amarillo proción, amarillo lucifer o cloruro de cobalto) se difunde a través de gran parte de la neurona, pero no atraviesa la barrera de la membrana celular.

Encontramos que cada interneurona gigante respondía de manera característica a soplos normalizados de aire dirigidos hacia los pelos cercales desde distintas direcciones. De las siete interneuronas gigantes situadas a cada lado del sistema nervioso (es decir, siete pares en conjunto), dos de cada lado respondían con igual fidelidad a los soplos de aire procedentes de cualquier dirección. En el sistema convencional de numeración, a estas dos interneuronas gigantes se las designa con los números 2 y 4. Otras dos de cada lado (los números 1 y 7) respondían de forma más enérgica al viento que golpeaba los cercos en el mismo sitio que ocupaban los axones de las interneuronas gigantes. Las tres interneuronas gigantes restantes de cada lado (números 3, 5 y 6) daban respuestas que eran más específicamente direccionales.

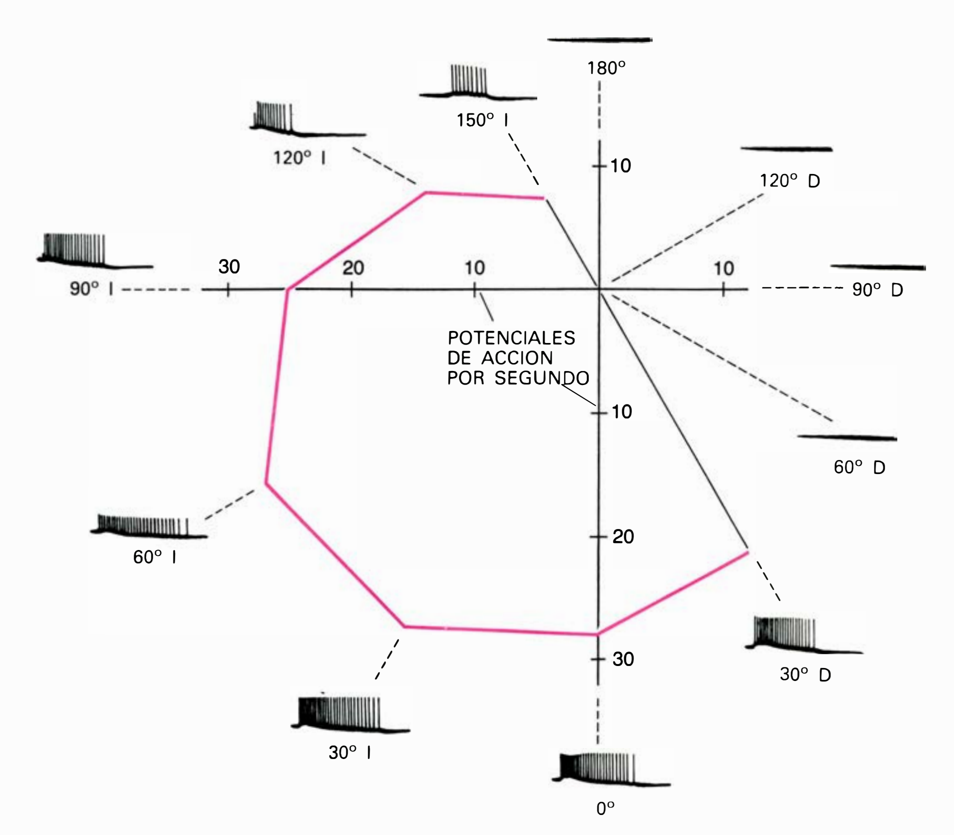
Ahora estamos empeñados en el examen de la pauta de conexión de determinadas columnas de células sensoriales con ciertas interneuronas gigantes al objeto de definir de qué manera se establecen las respuestas direccionales de cada interneurona gigante. En esa línea, hemos postulado que los dos pares de interneuronas gigantes que responden aproximadamente igual a todas las direcciones del viento deben recibir entradas excitadoras de varias columnas distintas de pelos cercales, si ha de explicarse su amplia respuesta direccional. En cam-

bio, los tres pares de interneuronas gigantes con respuestas direccionales bastante específicas han de recibir entradas excitadoras procedentes de una única columna de pelos.

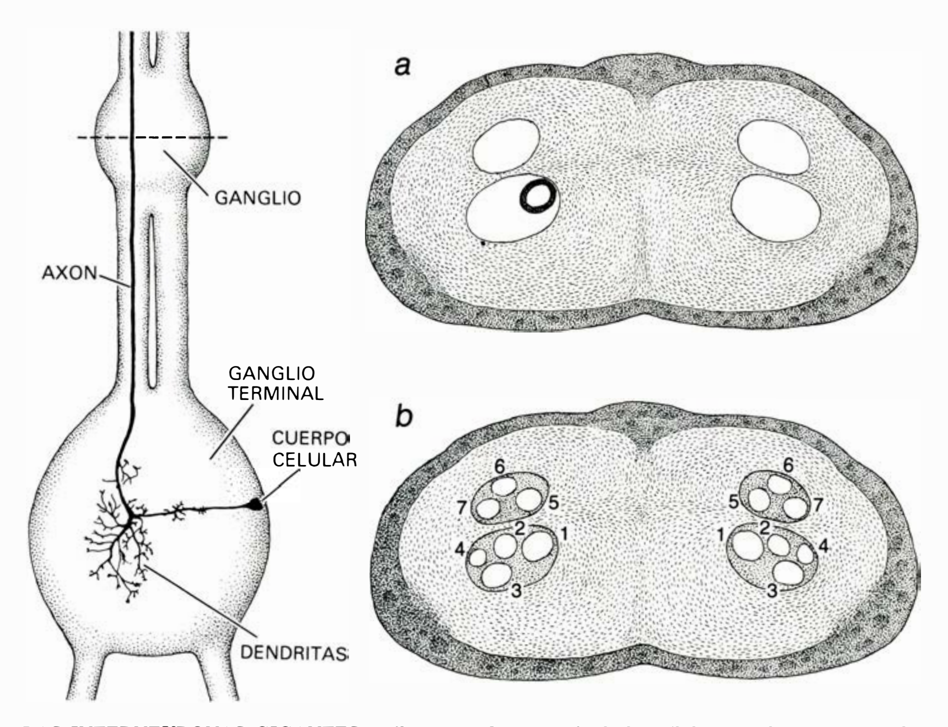
Apoyados en las respuestas observadas en las interneuronas gigantes, podría suponerse que, para determinar la dirección de un determinado soplo de aire, la cucaracha precisa basarse simplemente en la información que transmite uno de los tres pares de interneuronas gigantes, cuyas respuestas son las más claramente direccionales. Por ejemplo, si la interneurona gigante direccional del lado izquierdo denominada número 5 responde con su máximo número de potenciales de acción, el viento debe venir del cuadrante posterior, entre 30 y 60 grados a la izquierda del centro.

Sin embargo, la respuesta de las interneuronas gigantes no sólo depende de la dirección del viento, sino también de la velocidad instantánea del viento, de la que puede deducirse la aceleración. En consecuencia, no hay una relación fija entre el ritmo de disparo de una célula y la dirección del viento. Para obtener información direccional absoluta, la cucaracha debe "comparar" de alguna manera los ritmos de disparo de dos o más interneuronas gigantes. Puesto que todas las interneuronas gigantes tienen un ritmo de disparo característico para cada velocidad del viento, la comparación de los ritmos de disparo reales debe revelar diferencias relacionadas exclusivamente con la dirección del viento.

Aunque todavía no se sabe qué interneuronas gigantes son las más importantes a la hora de proporcionar esta información comparada y de evocar el comportamiento apropiado, una hipótesis atractiva sostiene que las unidades izquierda y derecha números 1, 2 y 3 intervienen en la respuesta de comportamiento inicial y especifican la dirección inicial de giro. (Puesto que la unidad número 2 tiene una respuesta omnidireccional, pudiera ser que sólo las números 1 y 3 especificaran la dirección.) Ocurre también que los axones de estos tres pares de interneuronas gigantes conducen potenciales de acción más rápidamente que ningún otro axón, lo que representa una importante ventaja para iniciar un rápido comportamiento de huida. Además, los potenciales de acción en respuesta a los soplos de aire se inician, en las interneuronas gigantes números 1, 2 y 3, unos cinco milisegundos antes que en cualesquiera de las otras cuatro.



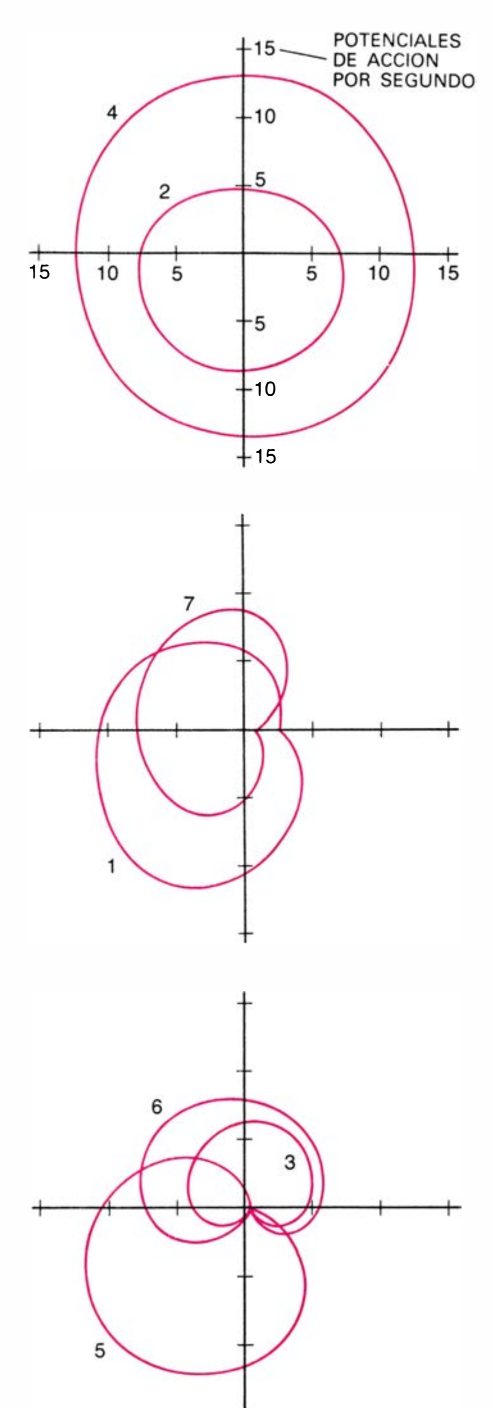
RESPUESTA DE UNA SOLA CELULA RECEPTORA del viento; puede apreciarse a través del registro intracelular desde el axón, o fibra larga, de la célula. Las respuestas de una célula del cerco izquierdo se dibujan en coordenadas polares para soplos de aire procedentes de 12 direcciones separadas 30 grados. Los soplos de aire procedentes de siete de las 12 direcciones generaron ráfagas de potenciales de acción, siendo máxima la respuesta evocada por los soplos localizados entre 30 y 60 grados hacia la izquierda. Se han representado aquí ejemplos promedio de los registros intracelulares para las siete direcciones.



LAS INTERNEURONAS GIGANTES reciben entradas a partir de las células cercales receptoras de viento y las reenvían a centros superiores del sistema nervioso de la cucaracha. La ilustración de la izquierda representa un cordón nervioso preparado *in toto*, en el que una de las interneuronas gigantes del insecto, la interneurona gigante número 1 del lado izquierdo, se ha hecho visible mediante tinción con el colorante cloruro de cobalto. El cuerpo celular se halla a la derecha del "árbol" de dendritas, que recibe la mayoría de las señales de entrada. Otras interneuronas gigantes tienen morfologías dendríticas distintas. El largo cable vertical es el axón celular, que transporta las propias señales de la célula. Las ilustraciones de detalle de la derecha, reproducidas de micrografías, muestran cortes de un ganglio que se trató de dos formas distintas. En *a* se destaca la interneurona gigante número 1 de la izquierda, teñida con el colorante amarillo proción. En *b* se muestran las 14 interneuronas según una tinción convencional.

Puesto que el comportamiento de huida puede iniciarse tan sólo 11 milisegundos después de que un soplo de aire particularmente acelerado alcance los receptores cercales de una cucaracha que camina lentamente, la ventaja de cinco milisegundos en el disparo parece eliminar como iniciadoras del comportamiento a todas las interneuronas gigantes que no sean las números 1, 2 y 3, al menos bajo estas especiales condiciones de latencia mínima.

En una extensión de esta hipótesis, después de haberse iniciado la acción de evitación, las interneuronas gigantes números 5, 6 y 7 podrían comandar la dirección de la respuesta de fuga. No se advierte aquí ningún papel específico de la interneurona gigante número 4, que, aunque se asemeja a la unidad número 2 en su omnidireccionalidad, no exhibe su elevada velocidad de conducción ni su capacidad de una rápida puesta en marcha.



RESPUESTAS DIRECCIONALES de siete interneuronas gigantes izquierdas a soplos de aire normalizados. Las interneuronas números 2 y 4 (*arriba*) responden casi exactamente igual a soplos de aire procedentes de cualquier dirección. Las interneuronas números 1 y 7 (*centro*) responden por lo general a soplos que alcanzan los receptores desde el lado izquierdo. Las interneuronas números 3, 5 y 6 (*abajo*) muestran respuestas que son algo más específicamente direccionales. Las siete interneuronas gigantes del lado derecho del cordón nervioso central de la cucaracha emiten respuestas que no son sino imágenes especulares de éstas.

Recientemente hemos dirigido nuestra atención hacia la plasticidad del sistema de huida: los cambios adaptativos que tienen lugar en el comportamiento y el funcionamiento neural de la cucaracha como resultado de la experiencia adquirida tras una alteración sensorial. En uno de nuestros primeros experimentos arrancamos el cerco izquierdo a todos los miembros de un grupo de cucarachas; observamos que los insectos giraban predominantemente a la izquierda, con independencia de cuál fuese la dirección del viento. Puesto que en nuestras colonias encontramos muchas cucarachas que carecían total o parcialmente de un cerco (probablemente arrancado en luchas con otras cucarachas), sospechamos que en las poblaciones naturales los insectos mutilados debían realizar frecuentemente giros inadecuados, casi siempre con resultados fatales, a menos que se desarrollaran de manera relativamente rápida mecanismos adaptativos para alterar sus giros.

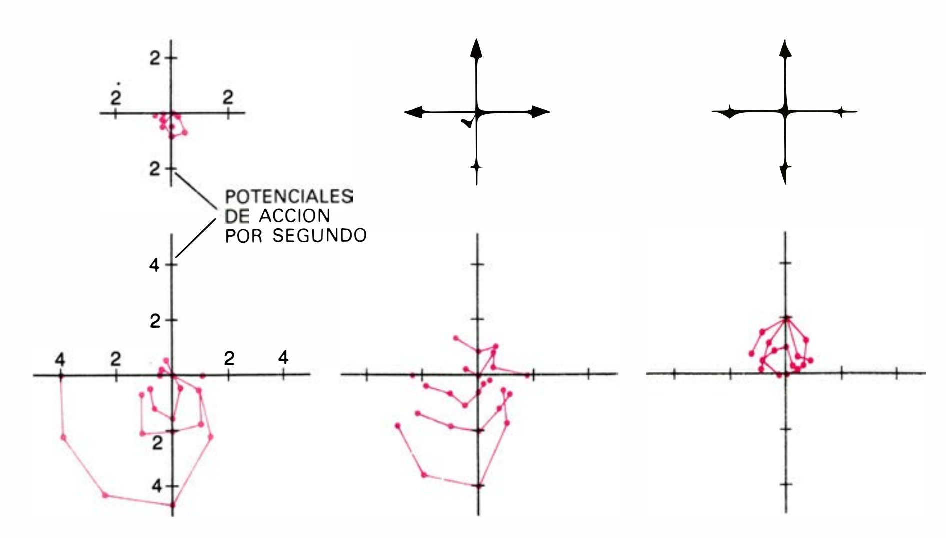
Noga Vardi, de nuestro laboratorio, ha probado la capacidad de las cucarachas con un cerco de menos para corregir sus giros inicialmente erróneos. Esta investigadora utilizó ninfas de cucaracha del último estadio, cuyo comportamiento de giro es normalmente tan preciso como el de los adultos. Cuando se las sometió a prueba un día después de la extirpación del cerco izquierdo, las ninfas solían girar hacia la izquierda en respuesta a soplos de aire procedentes de la izquierda. Sin embargo, treinta días más tarde, las ninfas impedidas se las apañaban bastante mejor, aunque no tan bien como los insectos normales. Además, no era preciso someter a prueba a las ninfas antes del trigésimo día para apreciar esta mejora. A lo largo del período de 30 días la mayoría de insectos no había regenerado aún una nueva yema cercal ni ningún nuevo pelo receptor de viento. Sin embargo, unos pocos individuos habían mudado y habían regenerado una nueva yema cercal con, todo lo más, unos cuantos pelos. Estas yemas se extirparon el día de su aparición.

Después de observar la corrección de comportamiento del trigésimo día, extirpamos el cerco derecho restante a un grupo de cucarachas; volvimos a probar el comportamiento del insecto al día siguiente. En muy raras ocasiones, los insectos que carecían de ambos cercos mostraron una levísima respuesta al viento de cualquier dirección. En este aspecto se comportaron exactamente igual que si se les acabara de extirpar ambos cercos. Llegamos a la conclusión de que, a lo largo del período de 30 días siguiente a la extirpación del cerco izquierdo, el sistema nervioso de la cucaracha aprovechaba la información direccional procedente de los pelos sensoriales del cerco derecho con el fin de efectuar un número significativamente mayor de vueltas correctas en comparación con las que efectuaba a las veinticuatro horas de la extirpación. Naturalmente, cabría esperar resultados afines si extirpáramos el cerco derecho de los insectos en experimentación.

¿Cómo puede explicarse esta mejora? Cuando se elimina el cerco izquierdo, muchas de las interneuronas gigantes del lado izquierdo dejan prácticamente de responder al viento de cualquier dirección. Sin embargo, las interneuronas gigantes del lado derecho continúan respondiendo casi normalmente, lo que significa que responden en alguna medida incluso al viento procedente de la izquierda. Evidentemente, las respuestas de las interneuronas gigantes del lado derecho al viento procedente de la izquierda son responsables de los giros (incorrectos) a la izquierda, del primer día. Hacia el trigésimo día, las interneuronas gigantes del lado izquierdo han recuperado entre una cuarta y una tercera parte de su número normal de potenciales de acción. Las unidades del lado derecho no muestran un cambio significativo durante el mismo período. En consecuencia, hay un restablecimiento parcial del ritmo normal de actividad de izquierda a derecha. Todas las interneuronas gigantes se hallan ahora dirigidas por receptores situados en el cerco derecho. (Si se cubre este cerco se bloquea toda respuesta al viento.)

Queda todavía mucho por aprender sobre la intensificación de la actividad que se registra en las interneuronas gigantes del lado izquierdo a lo largo del período de 30 días subsiguientes a la extirpación del cerco izquierdo.

Rodney K. Murphey y sus colaboradores, de la Universidad estatal de Nueva York en Albany, han apreciado una intensificación similar en la actividad de las interneuronas gigantes del lado izquierdo después de la extirpación del cerco izquierdo en grillos. Estos insectos tienen receptores cercales e interneuronas gigantes semejantes a los de las cucarachas. (Elizabeth Sherman, de nuestro laboratorio, ha descubierto que los grillos, como las cucarachas, muestran respuestas de giro a los soplos de aire y a las acometidas realizadas por sapos.) En sus estudios de grillos, Murphey y Rick Levine encontraron que buena parte de la actividad interneuronal intensificada procedía de una reducción en la fuerza de la inhibición sináptica ejercida sobre las interneuronas gigantes izquierdas desde el cerco derecho, sin que existiera una reducción concomitante en el nivel de excitación sináptica. No tenemos todavía pruebas de que este mecanismo, o algún otro, sea responsable de aquella intensificación de las respuestas de las interneuronas gigantes izquierdas que coincide con la mejora en el comportamiento de la cucaracha durante el período de 30 días.

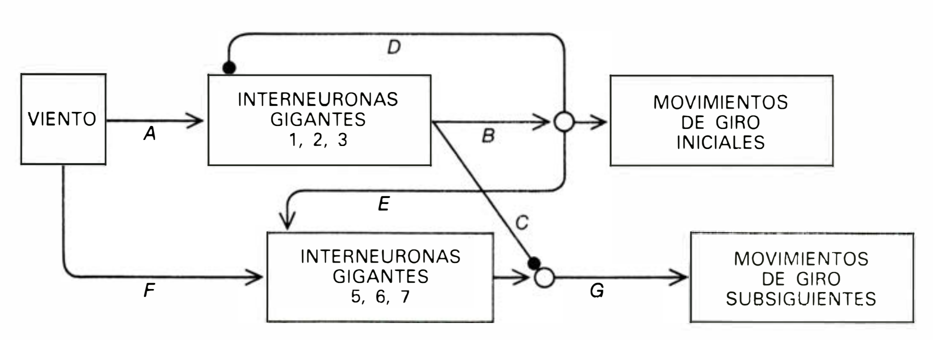


RECUPERACION DE LA ACTIVIDAD en las interneuronas gigantes, observada a los 30 días de la extirpación del cerco izquierdo de cucarachas ninfales. Los tres gráficos superiores indican las respuestas medias en varias pruebas de las interneuronas gigantes números 1, 2 y 3 un día después de haberse extirpado el cerco. En la mayoría de ninfas ensayadas no hubo respuesta a los soplos de aire, fuera cual fuese la dirección de procedencia, y de ahí la escasez de puntos. Los tres gráficos inferiores indican la recuperación que había tenido lugar cuando se sometió a prueba a los insectos 30 días más tarde. Las tres interneuronas exhiben pautas de respuesta que se parecen a las observadas en los insectos adultos normales, como se ha indicado en la ilustración precedente. (Las ilustraciones son de Tom Prentiss.)

Sea cual fuere el mecanismo celular que subyace a la recuperación de las interneuronas gigantes izquierdas, sigue planteada la cuestión de qué tipo de señal activa ese mecanismo cuando se extirpa un cerco. La eliminación de un cerco lleva a la degeneración del nervio cercal, y una posible señal sería algún subproducto químico de la degeneración. Sin embargo, los experimentos de comportamiento sugieren que la ausencia de actividad de las células sensoriales izquierdas, sin degeneración del nervio, es suficiente para iniciar el proceso que desemboca en la corrección del comportamiento. Por ejemplo, si cubrimos el cerco izquierdo con un adhesivo en lugar de extirparlo, se observa una pauta similar de vueltas incorrectas en el primer día y una corrección estadísticamente significativa hacia el trigésimo día. Un estudio al microscopio electrónico del nervio sensitivo y cierta cantidad de registros intracelulares procedentes de los axones sensoriales después de que el cerco hubiera permanecido cubierto durante 30 días indican que las células receptoras de viento no habían degenerado.

Hemos demostrado que la cantidad de actividad de los receptores de viento del cerco derecho influyen también en el grado de corrección del comportamiento. Pudimos deducirlo a partir de observaciones de cucarachas alojadas individualmente en pequeñas tazas de plástico, a lo largo del período de recuperación de 30 días. Las tazas protegían casi completamente a los insectos del viento ambiental. Además, como sea que los insectos tenían poco espacio para moverse, no había prácticamente viento autogenerado. Estos insectos confinados mostraron en un nivel significativo menos corrección que los que se mantuvieron en grandes jaulas protegidas. (Las cucarachas con dos cercos normales mantenidas en tazas durante 30 días conservaron su capacidad de dar la vuelta correctamente.) De hecho, las cucarachas mantenidas en las jaulas grandes no presentaban ninguna recuperación si su cerco izquierdo se había extirpado 30 días antes y su cerco derecho había permanecido cubierto durante el mismo período y después se destapaba para la prueba de comportamiento.

A partir de estas observaciones y de otras de tipo similar, llegamos a la conclusión de que la cantidad de actividad del nervio cercal derecho de la cucaracha era un determinante importante de la corrección del comportamiento después de la extirpación del cerco izquierdo. En realidad, se diría que lo que determinaba la extensión de la corrección del comportamiento era el grado de desequilibrio entre la actividad del nervio cercal izquierdo y la del nervio cercal derecho a lo largo del intervalo de 30 días. El sistema nervioso de la cucaracha parecía medir este desequilibrio y ajustar en consecuencia el grado de respuesta de las interneuronas gigantes. Sigue siendo un misterio cómo se logran la medición y el ajuste.



"CIRCUITO" DEL COMPORTAMIENTO DE HUIDA, según puede inferirse trazando las interacciones conocidas entre las células sensitivas, las interneuronas gigantes y otras neuronas en respuesta a soplos de aire. Los soplos que alcanzan los pelos cercales excitan a las interneuronas gigantes números 1, 2 y 3 a través de la ruta *A*, unos cinco milisegundos antes de excitar a las interneuronas números 5, 6 y 7 a través de la ruta *F*. Además, las interneuronas números 1, 2 y 3, una vez activadas, inhiben la salida de las interneuronas números 5, 6 y 7 a través de la ruta *C*. En un principio, por tanto, las unidades números 1, 2 y 3 ejercen probablemente el control dominante sobre los movimientos iniciales de giro. Una vez han dado comienzo los movimientos de giro, sin embargo, el circuito de retroalimentación inhibe a las unidades números 1, 2 y 3 a través de la ruta *D* y excita a las unidades números 5, 6, y 7 a través de la ruta *E*, de manera que los números 5, 6 y 7 pueden asumir el control dominante del comportamiento. Las interacciones fueron establecidas por Roy E. Ritzmann, Joanne Westin, Darryl L. Daley y el autor.

Formación de imágenes en los espejos

Las imágenes que proporcionan los espejos no corresponden siempre a una reflexión fidedigna. Estas imágenes pueden ser invertidas, trastrocadas en su lateralidad o distorsionadas

David Emil Thomas

La forma en que un espejo plano modifica (o no modifica) los objetos que refleja es algo que resulta familiar para todos nosotros. La imagen de una mano derecha aparece como si fuera una mano izquierda, pero su forma y tamaño permanecen inalterados. Sin embargo, tal como se pone de manifiesto al visitar determinadas barracas de feria, no todos los espejos dan lugar a este tipo de imágenes fidedignas. Platón observó que la imagen de un guante de la mano derecha, cuando se refleja en una pieza de metal curvada situada en el interior de un cilindro cóncavo, corresponde también a una mano derecha. En términos más generales, combinando adecuadamente ciertas superficies reflectoras de curvaturas diferentes, se pueden construir espejos capaces de cambiar la forma, el tamaño, la orientación y la lateralidad de los objetos que reflejan de manera tan espectacular como sorprendente.

El número de superficies especulares curvadas diferentes que podemos tomar en consideración es infinito, pero todas estas superficies se pueden obtener a partir de un número relativamente pequeño de elementos básicos. Cualquier región limitada de una superficie curva se puede describir en términos de su curvatura respecto a un par de ejes perpendiculares cuyo origen está situado en el centro de curvatura de la región en cuestión. Las direcciones de los ejes principales vienen determinadas por las direcciones para las que la curvatura de dicha superficie es máxima o mínima. La curvatura de cada eje puede ser convexa, cóncava o plana. Estos ejes nos permiten definir seis tipos básicos de superficies reflectoras: cada uno de estos tipos constituye un espejo simple cuya forma general viene especificada por dos curvaturas principales. Estas seis superficies básicas son el espejo plano, en el que los dos ejes principales son coplanarios; el espejo convexo, en el que ambos ejes son convexos; el espejo cóncavo, en el que ambos ejes son cóncavos; el cilindro convexo, en el que un eje es convexo y el otro plano; el cilindro cóncavo, en el que un eje es cóncavo y el otro plano y, finalmente, el espejo con forma de silla de montar, en el que uno de los ejes es convexo y el otro cóncavo.

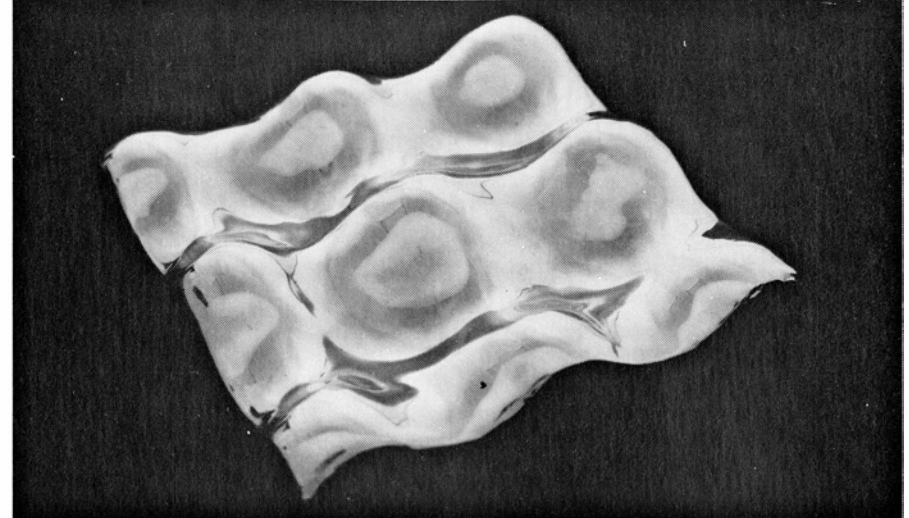
Durante estos últimos cinco años, he estado estudiando ciertas superficies reflectoras especiales tales como campanas y toroides; estas superficies se pueden construir a partir de los seis tipos básicos citados anteriormente. Las imágenes que proporcionan estos espejos compuestos son muy diversas, pero, en cualquier caso, he llegado al descubrimiento de que tales imágenes presentan ciertas regularidades llamativas. Esos rasgos fundamentales pueden atribuirse a los elementos básicos que constituyen dichos espejos. Es posible predecir la aparición de unas determinadas imágenes reflejadas incluso en espejos compuestos muy complejos. En consecuencia, con el fin de poder estudiar los espejos compuestos y las imágenes que proporcionan, conviene repasar previamente las propiedades reflectoras de los seis espejos básicos.

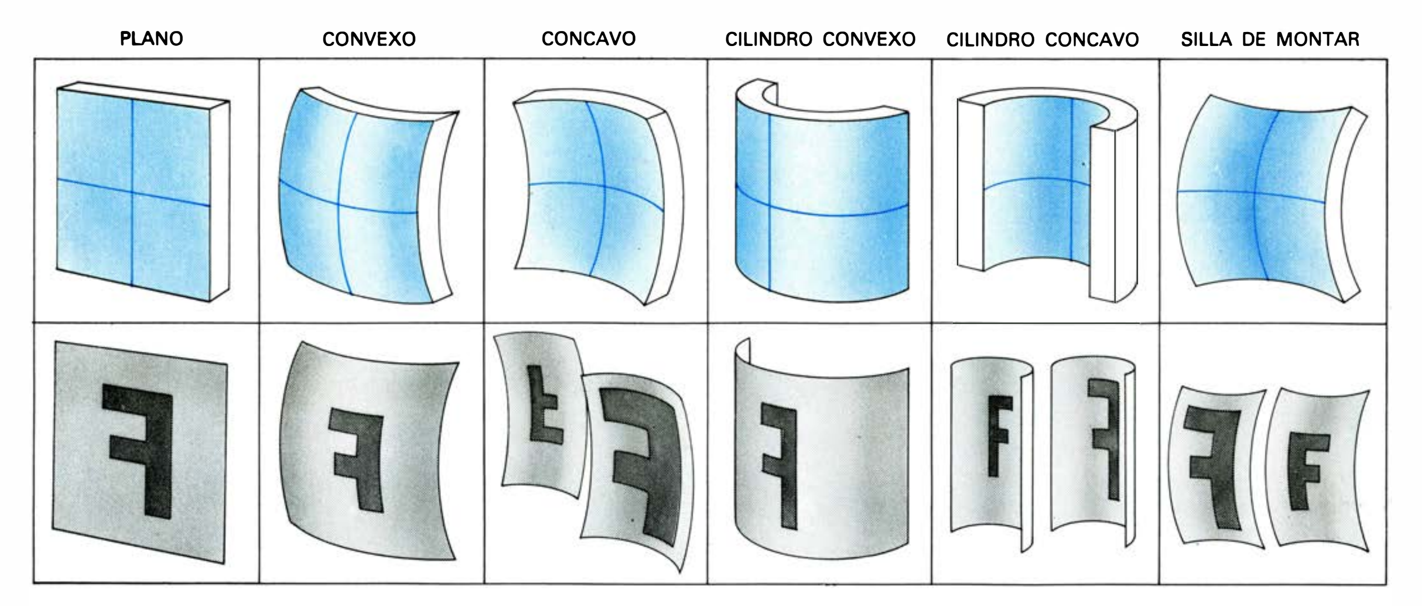
Para empezar, vamos a tomar en consideración una curiosa e importante cuestión en relación con los espejos planos. ¿Por qué estos espejos cambian la derecha con la izquierda y no la parte de arriba con la de abajo? En otras palabras, ¿por qué cuando se coloca un libro ante un espejo las letras aparecen invertidas y la secuencia de estas letras va de derecha a izquierda, en tanto que el orden de las líneas se mantiene inalterado desde la parte superior hasta la parte inferior de las páginas? Semejante paradoja está relacionada con la confusión que surge entre la lateralidad de una imagen (izquierda y derecha) y su orientación. En realidad, para un observador que se coloque frente a un espejo plano, las direcciones derecha e izquierda no cambian más de lo que lo hacen las direcciones arriba y abajo. Este hecho se puede comprobar comparando la esfera de un reloj con la imagen que forma un espejo plano.

La conocida geometría de una esfera de reloj constituye un objeto de referencia muy apropiado para el estudio de las relaciones objeto-imagen en los diferentes espejos curvos. En mis propias investigaciones sobre este tema, he utilizado una esfera de reloj cuyos números están grabados sobre un plástico transparente. Un observador que mantenga este reloj frente a un espejo será capaz de ver no sólo la imagen que proporciona el espejo, sino también el objeto real, al menos desde atrás. A través del plástico transparente, el observador ve los números del reloj objeto invertidos y con las saetas moviéndose en sentido inverso al usual, al mismo tiempo que observa las imágenes que forman unos y otros.

Esta esfera transparente de reloj es particularmente útil para nuestro estudio, dado que los números permiten definir las direcciones laterales del plano tangente al espejo en su centro de curvatura. Por ejemplo, las 11:00 identifican la dirección que va desde el centro del reloj hasta el número 11; de forma análoga, las 5:30 identifican la dirección que va del centro del reloj al punto situado entre los números 5 y 6. (A lo largo de todo este examen, supondremos que el observador y el objeto se hallan próximos al eje del espejo, es decir, a la línea que pasa por

IMAGENES MULTIPLES de un objeto individual, en este caso una letra *F*, formadas por un "espejo de matriz". La superficie de este espejo presenta una curvatura completa, tal como se pone de manifiesto en estas figuras, que muestran dicha superficie vista desde arriba y oblicuamente. El espejo de matriz cuya superficie se halla suavemente unida, está constituido por una serie de regiones convexas, cóncavas y con forma de silla de montar, cada una de las cuales refleja una imagen diferente del objeto *F* considerado. Como consecuencia de las diferentes propiedades reflectoras de las distintas regiones que forman dicho espejo, la lateralidad de algunas imágenes está trastrocada, lo que no ocurre con la de otras. El espejo de matriz representado aquí es uno de los infinitos tipos de espejos compuestos que pueden crearse a partir de seis clases principales de ellos.





ESPEJOS CURVADOS BASICOS, a partir de los cuales se pueden construir todos los espejos compuestos. Podemos caracterizarlos en función de un par de ejes perpendiculares (*color*) que atraviesan el centro de curvatura de cada espejo. Las direcciones de los ejes principales vienen determinadas por las direcciones de máxima y mínima curvatura sobre las correspondientes superficies; la forma de cada eje puede ser convexa, cóncava o plana. Estos ejes permiten definir seis tipos de superficies reflectoras: el espejo plano, en el que ambas curvaturas son planas; el espejo convexo, en el que ambas curvaturas son convexas; el espejo cóncavo, en el que ambas curvaturas son cóncavas; el cilindro convexo, en el que una de las curvaturas es convexa y la otra plana; el cilindro cóncavo, en el que una curvatura es cóncava y la otra plana y la silla de montar, en la que una curvatura es cóncava y la otra convexa. Las propiedades de estos espejos, por lo que respecta a la formación de imágenes, se ponen de manifiesto en la parte inferior de estas figuras, utilizando una letra *F* como objeto muestra. En algunos casos, un espejo determinado puede dar lugar a imágenes diferentes para circunstancias diferentes. Se supone que observador y espejo están situados sobre el eje del espejo: sobre una línea perpendicular que pasa por su centro de curvatura.

el centro de curvatura del espejo y es perpendicular al plano tangente en este punto.)

Este sistema de coordenadas proporciona un medio de describir explícitamente la falsa diferencia que existe entre lateralidad y orientación. Las saetas que vemos en la imagen que proporciona un espejo cuando situamos frente a él una esfera de reloj transparente van pasando de un número a otro en un orden contrario al usual, es decir, su movimiento coincide con el que observamos a través de la parte de atrás de este mismo reloj. Cuando comparamos el objeto real con su imagen, "frente a frente", se pone de manifiesto que ni la dirección 9:00 (derecha), ni la 3:00 (izquierda), ni la 12:00 (arriba), ni la 6:00 (abajo), cambian en el espejo plano. Debido a que la orientación no cambia para ninguna dirección del plano del espejo, en la imagen de una esfera de reloj las saetas recorren los sucesivos números en una secuencia contraria a la usual. Pero los espejos planos (y todos los demás espejos) invierten las direcciones hacia adelante y hacia atrás: mientras que la esfera del reloj real se encuentra de espaldas al observador, su imagen en el espejo aparece de frente al observador. Esta simple inversión delante-detrás crea una imagen en la que las posiciones de la derecha y la izquierda están invertidas. Por la misma razón, la imagen de un guante de la mano derecha corresponde a la mano izquierda. Al objeto de evitar confusiones, en adelante utilizaré el término "invertir" para referirme a los cambios en la orientación de una imagen y el término "trastrocar" para describir los cambios en la lateralidad.

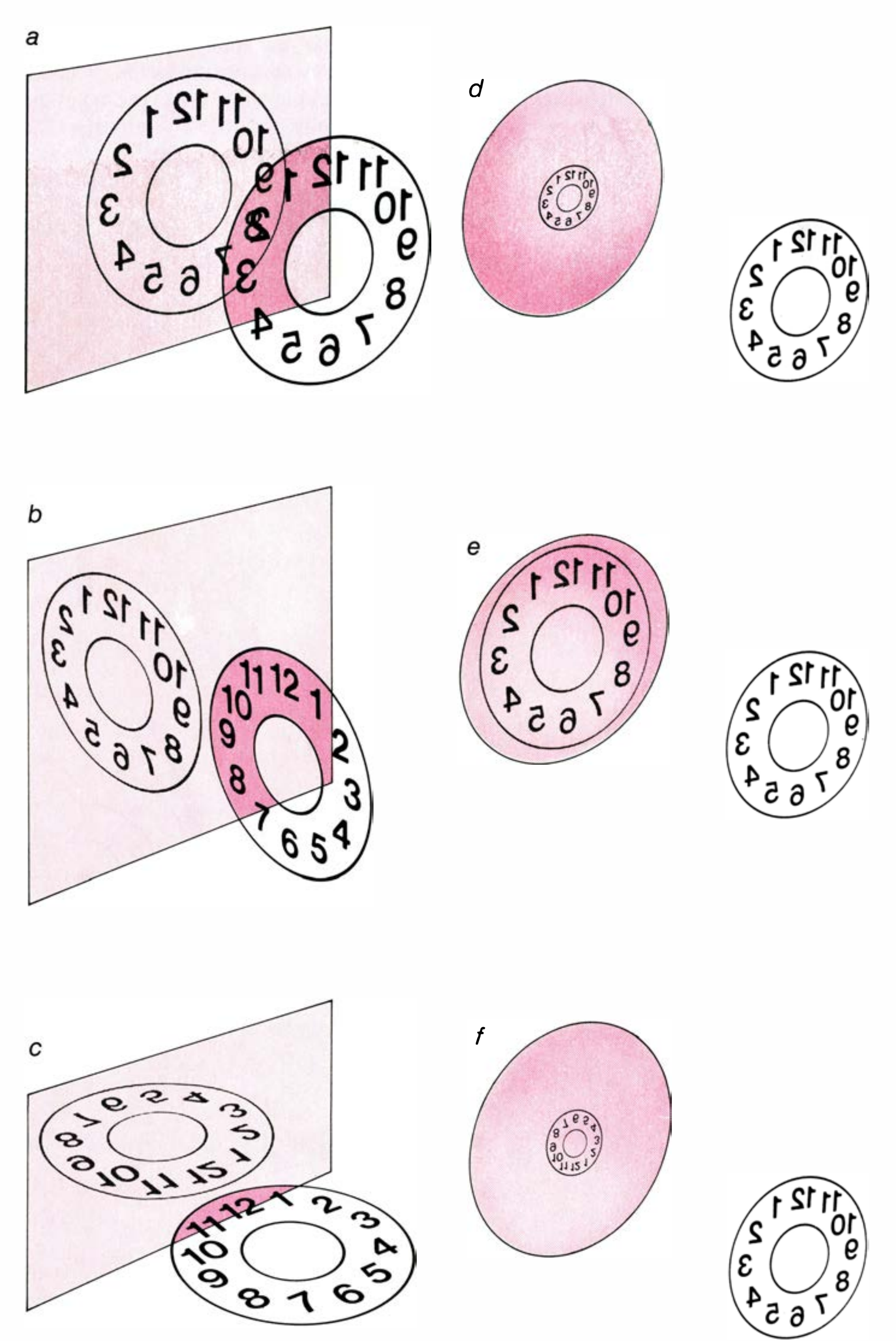
La inversión delante-detrás se pone de manifiesto al colocar un objeto en una posición paralela a la superficie de un espejo plano. Sin embargo, cuando colocamos la esfera de un reloj en un plano perpendicular al espejo, esta misma transformación conduce a un resultado diferente. Si la esfera del reloj se sitúa en un plano perpendicular al espejo y en posición vertical, la inversión delante-detrás cambia el 3 por el 9 (así como el 1 por el 11 y el 2 por el 10, entre otros), de forma tal que la derecha y la izquierda parecen estar invertidas. De modo análogo, si el reloj se mantiene en un plano perpendicular al espejo aunque en una posición horizontal, la inversión delante-detrás cambia el 6 por el 12 (así como el 1 por el 5 y el 11 por el 7, entre otros), de forma tal que el espejo parece invertir el arriba y el abajo.

La imagen que proporciona un espejo plano se conoce tradicionalmente con el nombre de imagen virtual; se trata de una imagen que parece estar detrás de la superficie del espejo. En otras palabras, las ondas luminosas reflejadas que forman las imágenes en un espejo plano parecen emanar de un punto situado detrás del espejo. La ley de reflexión de la óptica establece que el ángulo de reflexión de un rayo luminoso debe ser igual al ángulo de incidencia. Como consecuencia de esta ley, la imagen de un objeto en un espejo plano parece estar tan alejada por detrás del espejo como lo está el objeto por delante de él. La imagen de un objeto situado a la distancia de un metro frente a un espejo plano tiene el mismo tamaño, forma y orientación (pero con la derecha e izquierda trastrocadas) que un objeto situado a un metro detrás del espejo.

En el caso de espejos convexos podemos hacer unas consideraciones prácticamente idénticas a las anteriores. Los espejos convexos también dan lugar a imágenes virtuales. Al igual que los espejos planos, los espejos convexos invierten la orientación únicamente para la dirección delante-detrás y no para cualquier dirección lateral; en consecuencia, estos espejos reflejan imágenes verticales y trastrocadas. Sin embargo, dado que su superficie presenta una curvatura convexa, las ondas luminosas reflejadas parecen emerger de un punto que no está tan lejos por detrás del espejo como el punto en el que parecía estar situada la imagen de un espejo plano. En virtud de este hecho, la imagen de un espejo convexo resulta ser menor que el de su homóloga en un espejo plano.

Aquella dirección lateral para la cual un espejo curvo es plano o convexo se puede denominar dirección de no inversión. Por tanto, ¿qué entendemos por dirección de inversión? Para contestar a esta pregunta, vamos a ver lo que le sucede a un frente de ondas luminoso cuando se refleja en la superficie de un espejo curvo. El frente de ondas incidente es esférico y divergente. Tras la reflexión, el frente de ondas puede ser esférico o asimétrico y, por otra parte, puede ser divergente (posiblemente en una cuantía diferente de lo que lo era el frente de ondas incidente) o convergente, es decir, las partes que lo constituyen se pueden ir alejando o aproximando mutuamente. Es el paso de una parte de la onda a través de otra lo que da lugar a una inversión lateral. Este cruce tiene lugar en el punto donde el espejo focaliza las ondas luminosas incidentes para engendrar lo que se conoce como imagen real de un objeto. Las direcciones para las que el espejo hace que ciertas partes de una onda converjan determinan qué direcciones laterales se están invirtiendo en la reflexión. Si la orientación se invierte lateralmente a lo largo de los dos ejes principales de un espejo curvo básico, se forma una imagen con la lateralidad trastrocada. Si la orientación se invierte lateralmente a lo largo de un único eje lateral, se forma una imagen no trastrocada. Las inversiones laterales se presentan acompañadas de la inversión delante-detrás, que es común para todos los espejos. En general, si el número de inversiones (incluyendo la inversión delante-detrás) es impar, la imagen presenta la izquierda y la derecha trastrocadas; si el número de inversiones es par, la derecha y la izquierda no resultan trastrocadas.

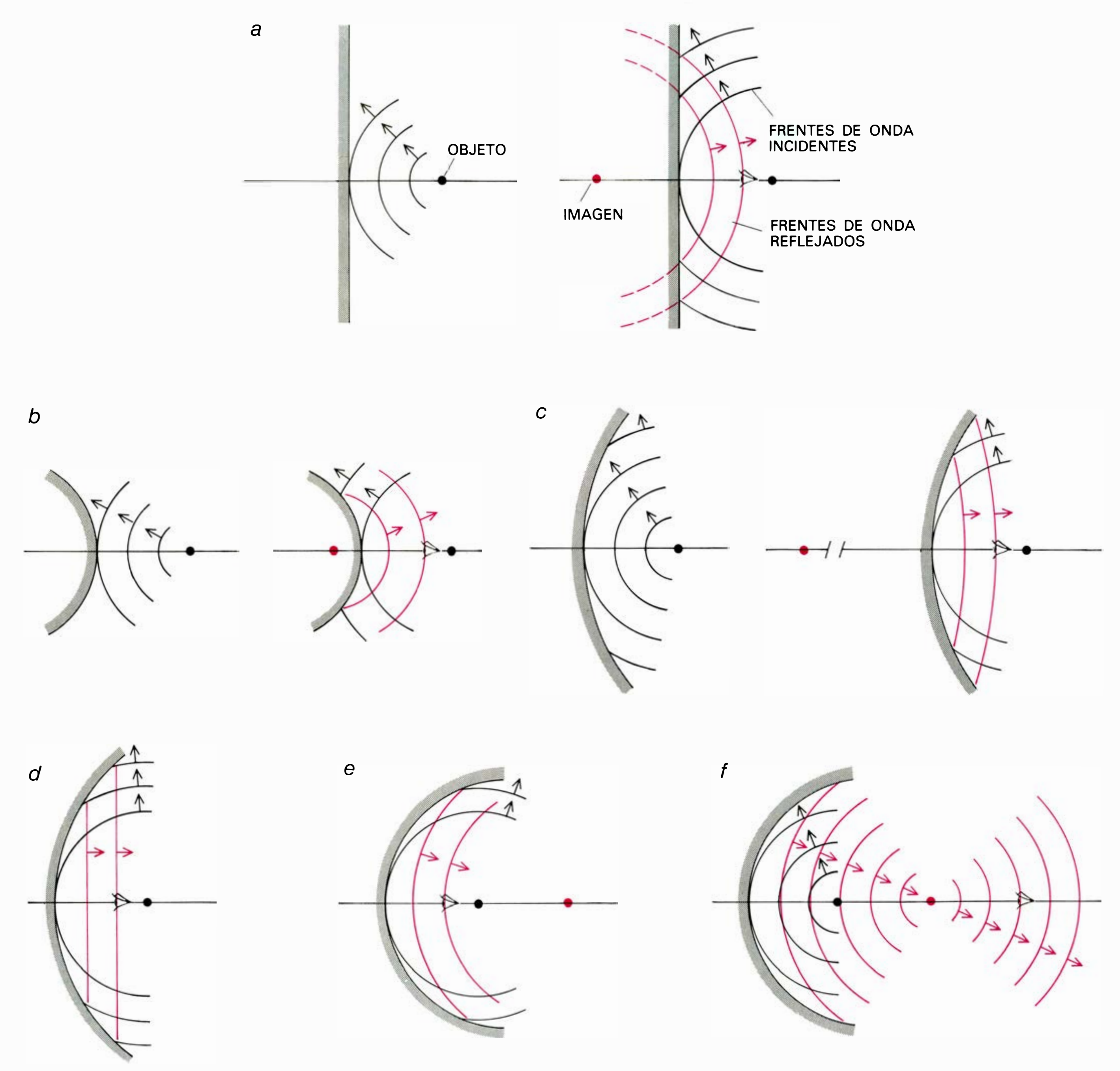
La posición del observador constituye un factor importante en la formación de imágenes en los espejos. Un observador puede ver una imagen real e invertida únicamente en el caso de que la luz reflejada por el espejo se focalice sobre un punto situado entre él y la superficie del espejo; en otras palabras, el observador sólo debe interceptar los frentes de onda después de que hayan convergido y no antes. Los espejos planos y convexos, que no producen la convergencia en los frentes de onda que inciden sobre ellos, no pueden dar lugar a inversiones laterales. Por otra parte, cuando un observado se sitúa muy cerca de un espejo curvo, no importa cuáles sean las propiedades de inversión del espejo en



ESFERA DE UN RELOJ CON UN FONDO TRANSPARENTE. Dicho dispositivo nos permite definir las distintas direcciones laterales en el plano tangente a un espejo por su centro de curvatura. Por ejemplo, las 2:00 identifican la dirección lateral que va desde el centro del reloj hasta el número 2 y las 10:30 identifican la dirección que va desde el centro del reloj hasta el punto medio situado entre los números 10 y 11. Un observador que sostuviera verticalmente dicho reloj frente a un espejo sería capaz de ver los números reales (con su lateralidad trastrocada y ordenados según una secuencia contraria al sentido en que avanzan las agujas de un reloj) al mismo tiempo que las imágenes que se forman en el espejo. Cuando este reloj se mantiene frente a un espejo plano y se compara "cara a cara" el reloj real con el reloj imagen (*a*), resulta evidente que el espejo plano no invierte la orientación de las direcciones laterales: el 2 aparece en la dirección 2:00 de la imagen, el 10 en la dirección 10:00 y así sucesivamente. Ni la dirección 12:00 (arriba), ni la 6:00 (abajo), ni la 9:00 (derecha), ni la 3:00 (izquierda) se han invertido. Sin embargo, aun cuando el espejo plano no produce inversiones laterales, la lateralidad de la imagen se ha trastrocado. Por otra parte, un espejo plano, como cualquier otro espejo, invierte la orientación a lo largo del eje antero-posterior, de forma tal que un observador mira al espejo para ver los objetos situados frente a él. Si la esfera del reloj se mantiene en un plano perpendicular a la superficie del espejo, se produce una inversión tanto de la derecha y la izquierda (*b*) como de arriba y abajo (*c*). Un espejo convexo (*d*) no invierte la orientación de ninguna dirección lateral y, por consiguiente, da lugar a una imagen derecha y con la lateralidad trastrocada. Para un espejo cóncavo, y en el caso de que tanto el observador como el objeto estén suficientemente próximos a la superficie del espejo (*e*), ninguna dirección lateral invierte su orientación y, por tanto, se forma una imagen derecha y trastrocada. Sin embargo, en el caso de que el observador y el objeto están alejados del espejo cóncavo (*f*), todas las direcciones laterales invierte su orientación; como consecuencia, en dicha circunstancia se forma una imagen que es invertida y trastrocada.

cuestión, la imagen resultante es similar a la que produciría un espejo plano. Los frentes de onda no disponen de espacio suficiente para converger antes de que alcancen el ojo del observador y, por tanto, la imagen es vertical, trastrocada y (en comparación con la imagen de un espejo plano) ni muy aumentada ni muy reducida. Cuando es el objeto (y no el observador) el que se sitúa muy próximo a la superficie del espejo, se forma también una imagen similar a la imagen que proporciona un espejo plano. Sin embargo, la explicación de este caso es muy diferente: cuando el objeto se coloca muy cerca de un espejo, la sección de superficie reflectora que interviene realmente en la formación de la imagen se puede considerar prácticamente plana.

Un espejo cóncavo puede dar lugar

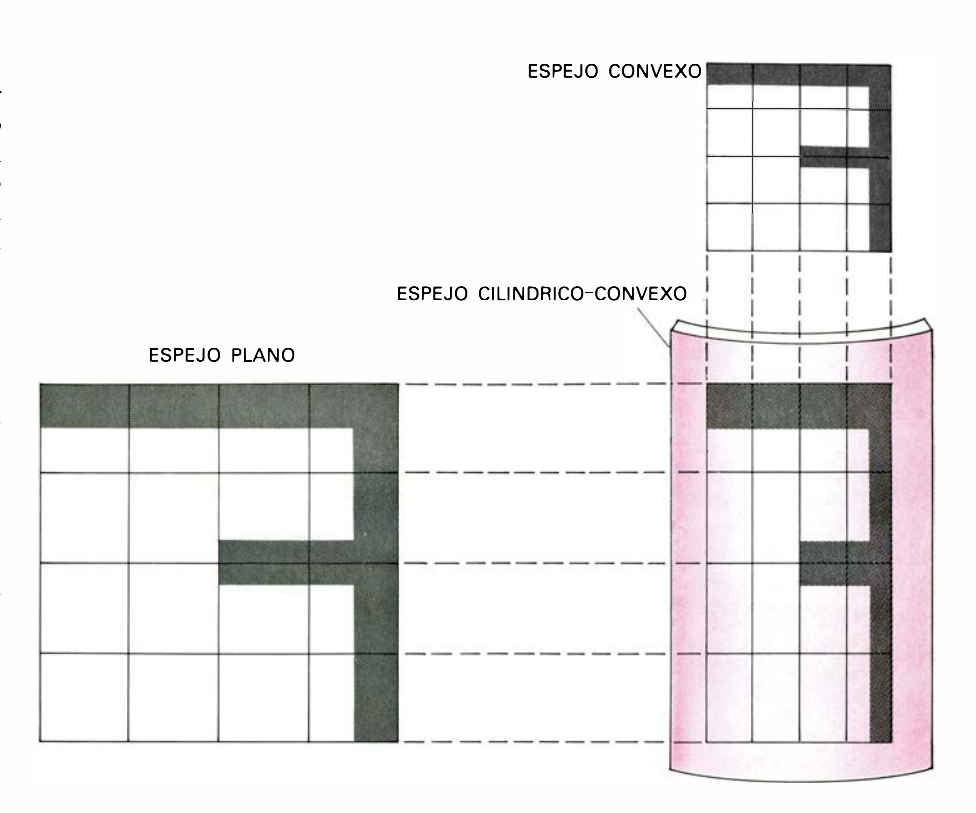


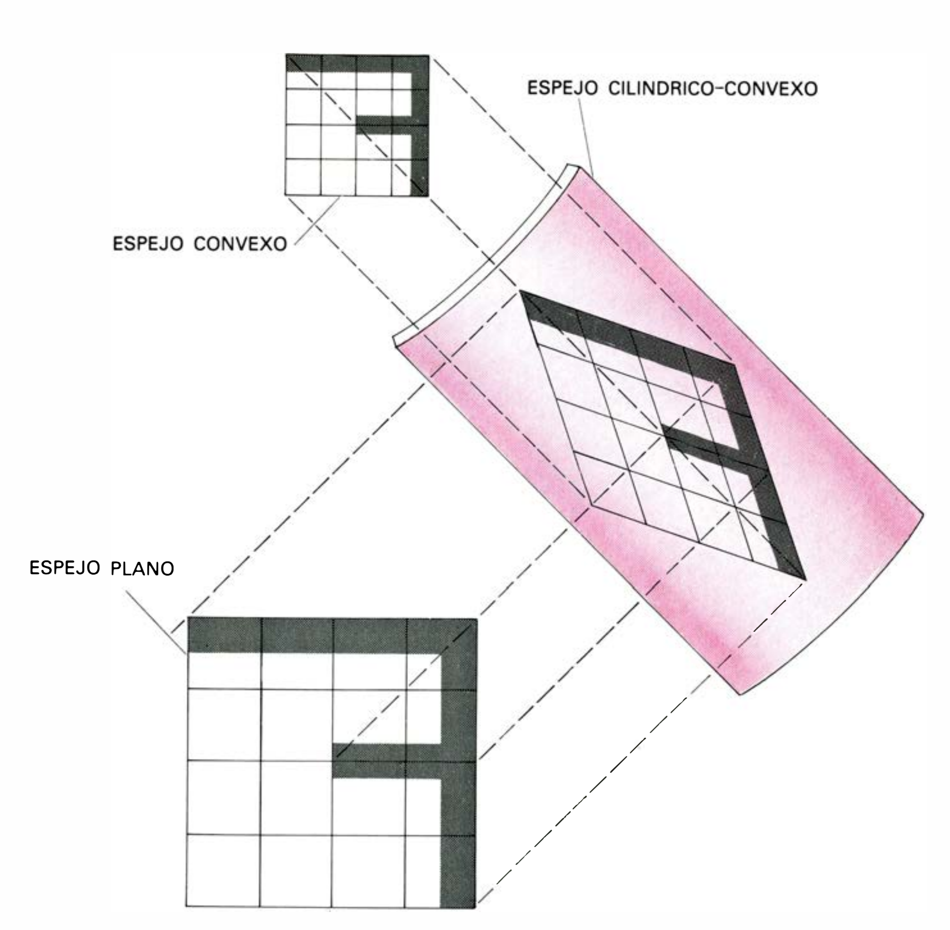
TAMAÑO, ORIENTACION Y LATERALIDAD de la imagen producida por un espejo. Tales propiedades vienen determinadas por la forma en que los frentes de onda se transforman en la reflexión. Los frentes de ondas que emergen del objeto (*gris*) son esféricos y divergentes. Los frentes de ondas reflejados (*color*) pueden ser divergentes o convergentes. En el caso de un espejo plano (*a*), los frentes de ondas reflejados parecen divergir a partir de un punto situado detrás de la superficie del espejo, en tanto que el objeto (*punto negro*) está situado delante de él. El observador verá una imagen derecha y trastrocada (*punto coloreado*), cuyo tamaño no es mayor ni menor que el del objeto real. En el caso de un espejo convexo (*b*), la situación es similar, aunque los frentes de ondas reflejados parecen divergir a partir de un punto situado un poco por detrás de la superficie del espejo y, como resultado, la imagen es más pequeña que el objeto. En el caso de un espejo cóncavo se pueden presentar toda una serie de posibilidades diferentes. Cuando el objeto se sitúa entre el espejo y su punto focal (el punto en el que se focalizan los rayos paralelos al eje del espejo), los frentes de ondas reflejados parecen divergir de un punto situado mucho más detrás del espejo de lo que lo estaría la imagen de un espejo plano (*c*). Se forma entonces una imagen trastrocada cuyo tamaño es mayor que el de la que se formaría en un espejo plano. Cuando el objeto está situado en el punto focal (*d*), los frentes de onda reflejados son planos y el observador debe focalizarlos, creando una imagen derecha cuyo tamaño es mayor que el que tendría la imagen formada por un espejo plano. Cuando el objeto está situado más allá del punto focal, los frentes de onda reflejados convergen delante del espejo. Un observador situado entre el espejo y el punto focal deberá focalizar los frentes de onda antes de que converjan, creando una imagen derecha, trastrocada y de mayor tamaño (*e*). Si el observador se sitúa más allá del punto focal, interceptará los frentes de onda después de que hayan convergido; la imagen será trastrocada e invertida (*f*). La posición en que se encuentra el observador determinará si la imagen es mayor, menor o igual que la obtenida con un espejo plano.

al menos a tres tipos diferentes de imágenes. Ante todo, se puede originar una imagen real o virtual, dependiendo de la posición del objeto respecto al foco del espejo, es decir, respecto al punto en que los frentes de onda planos convergen o parecen divergir. Si el objeto está situado entre el foco y el espejo, los frentes de onda divergen, partiendo aparentemente de un punto situado detrás del espejo y, por consiguiente, la imagen resultante es virtual. Cuando el objeto se coloca más allá del foco del espejo, los frentes de onda reflejados convergen a lo largo de todas las direcciones laterales y se focalizan en un punto situado en el mismo lado en que se encuentra el observador respecto al espejo. En este caso, la imagen es real.

En virtud de cuanto antecede, cuando un objeto está situado entre un espejo cóncavo y su foco, el espejo proporciona una imagen virtual para cualquier distancia entre el observador y el espejo. La imagen es derecha y trastrocada y resulta ser de mayor tamaño que la que proporcionaría un espejo plano. Cuando el objeto está más allá del foco del espejo, existen varias posibilidades. Por ejemplo, si el observador está situado entre la superficie del espejo y el punto sobre el que se focaliza la imagen real, intercepta los frentes de onda convergentes antes de que dichos frentes converjan realmente. En este caso, es el observador el que debe focalizar los frentes de onda y el resultado es una imagen borrosa, derecha, trastrocada y mayor que la que proporcionaría un espejo plano.

Cuando el observador está situado más allá de la posición que corresponde a una imagen real, el espejo cóncavo introduce inversiones en todas las direcciones laterales. En otras palabras, cuando tanto el observador como el objeto se hallan suficientemente lejos de la superficie del espejo, la esfera de un reloj transparente se refleja de forma tal que la imagen de cada número aparece desplazada 180 grados a lo largo de la esfera del reloj (es decir, un intervalo de seis horas). El número 1 aparece en las 7:00 y el 7 en la 1:00, el 2 aparece en las 8:00 y el 8 en las 2:00, y así sucesivamente. Puesto que las direcciones 12:00 (arriba) y 6:00 (abajo) están invertidas, la imagen es invertida. Las posiciones del objeto y del observador determinan si la imagen resulta aumentada, reducida o del mismo tamaño en comparación con su contrapartida en un espejo plano. Puesto que





EL CILINDRO CONVEXO se comporta como un espejo plano a lo largo de un eje lateral y como un espejo convexo a lo largo del otro eje lateral. Proyectando las imágenes obtenidas con estos espejos simples, se puede predecir el aspecto que presentará la imagen asimétrica que se obtiene en un cilindro convexo. La imagen en cuestión es derecha, trastrocada y de menor tamaño en relación con el eje de curvatura convexo. Para analizar la clase de imagen que se forma cuando el cilindro convexo se inclina hacia la izquierda o hacia la derecha, se puede utilizar este mismo tipo de construcción geométrica.

la orientación se invierte para todas las direcciones laterales, se invierte para ambos ejes principales de curvatura; estas dos inversiones, junto con la inversión delante-detrás, representan un número impar y, por consiguiente, la imagen no sólo es invertida sino trastrocada.

Con un espejo cóncavo y un objeto situado más allá del punto focal es posible conseguir que el observador y la imagen que proporciona el espejo estén situados en el mismo punto del eje del espejo. En este caso, el resultado es una imagen enormemente aumentada: una imagen irreconocible que "llena" toda la superficie del espejo. Cualquier espejo curvo capaz de formar una imagen real puede dar lugar a una imagen de este tipo.

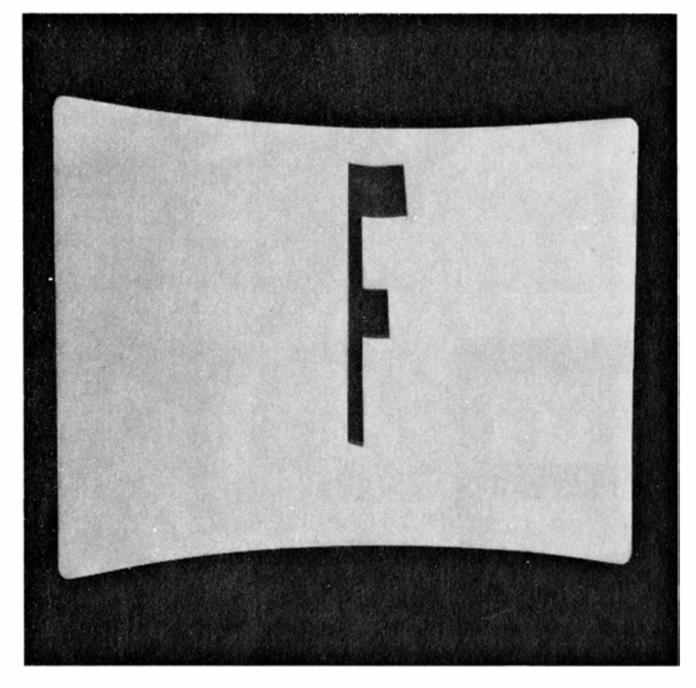
Los espejos curvos cuya sección no es esférica proporcionan imágenes de naturaleza asimétrica. Por ejemplo, las imágenes de los espejos convexos y cóncavos de forma oblonga se muestran un tanto alargadas o achatadas en determinadas direcciones laterales, en comparación con las que se obtendrían al utilizar espejos planos o esféricos.

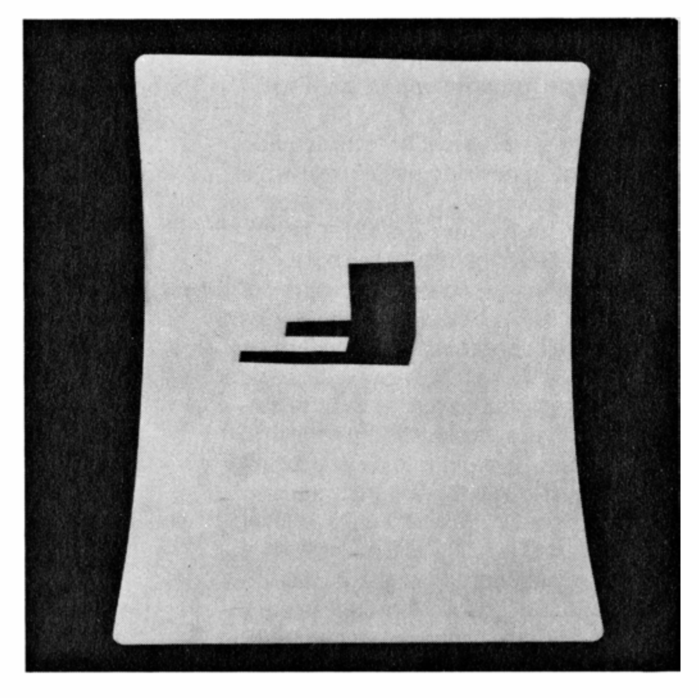
Cuando se conoce la curvatura de un espejo no simétrico respecto a los dos ejes principales, se puede predecir el aspecto general que presentarán las correspondientes imágenes. Vamos a considerar un espejo básico cuya forma corresponde a la de un cilindro convexo. Este espejo se comporta prácticamente como un espejo convexo para todas aquellas direcciones que son paralelas a la curvatura convexa y como un espejo plano para todas aquellas direcciones que son paralelas a la curvatura plana; las imágenes que engendra este espejo son de tamaño reducido respecto al primero de dichos ejes y del mismo tamaño respecto al segundo. Aplicando las leyes de la geometría plana y tomando como punto de partida las imágenes que se obtienen en un espejo plano y en un espejo convexo, se puede predecir con bastante aproximación la imagen que proporciona un cilindro convexo.

Este tipo de predicciones se puede aplicar también al caso de los cilindros cóncavos, un espejo básico que se comporta aproximadamente como un espejo cóncavo a lo largo de una de las curvaturas principales y como un espejo plano a lo largo de la otra curvatura. Si tanto el observador como el espejo están suficientemente cerca de la superficie del cilindro cóncavo (a una distancia tal para la que un espejo cóncavo daría lugar a una imagen derecha, trastrocada y de mayor tamaño que el objeto), la imagen reflejada por dicho cilindro será derecha, trastrocada y de mayor tamaño que el objeto. Por el contrario, si el observador y el objeto estuvieran lejos del espejo (lo bastante lejos como para que al utilizar un espejo cóncavo simple la imagen resultante fuera invertida y trastrocada), el resultado diferiría.

Los frentes de onda que se reflejan en la superficie de un cilindro cóncavo no sufren una inversión en todas las direcciones laterales, tal como sucede cuando se produce una reflexión en un espejo cóncavo, sino que tan sólo se invierten lateralmente a lo largo de la curvatura cóncava. Dado qué la imagen resultante presenta un total de dos inversiones (una lateral y la del tipo delante-detrás), no está trastrocada y, por consiguiente, la imagen de una mano izquierda es una mano izquierda. Dependiendo de las posiciones relativas del objeto y del observador, la imagen puede resultar aumentada, reducida o del mismo tamaño respecto a las direcciones paralelas a la curvatura cóncava. Por otra parte, la orientación lateral de la curvatura cóncava determina la orientación lateral de la imagen del espejo; de aquí que girando adecuadamente el eje del espejo se pueda dar cualquier orientación a la imagen resultante. Cuando el eje de la curvatura cóncava está situado en posición vertical, la imagen no trastrocada aparece con la parte superior hacia abajo.

Los frentes de onda que emanan de uno de los puntos de un objeto son esféricos hasta que alcanzan la superficie reflectora de un cilindro cóncavo. Tras sufrir una reflexión en esta superficie, dichos frentes de onda se hacen más y más asimétricos hasta que, finalmente, llegan a un foco. Sin embargo, dado que los frentes de onda convergentes no son simétricos, no se cruzan





EL CILINDRO CONCAVO se comporta como un espejo cóncavo a lo largo de un eje y como un espejo plano a lo largo del otro eje, de suerte que, cuando el observador y el objeto están suficientemente alejados del espejo, la orientación se invierte únicamente a lo largo de uno de los ejes y se forma una imagen no trastrocada. Cuando el eje de curvatura cóncava es vertical (*derecha*), se invierten las direcciones de arriba y abajo y, por tanto, se obtiene una imagen no trastrocada e invertida. Las posiciones del objeto y del observador determinan que el tamaño de la imagen sea mayor o menor.





ESPEJO CON FORMA DE SILLA DE MONTAR. Se comporta como un espejo convexo a lo largo de uno de los ejes y como un espejo cóncavo a lo largo del otro eje. Cuando el observador y el objeto están alejados, la superficie de este espejo invierte la orientación de aquellas direcciones que pasan por el centro de un valle. La orientación de la imagen viene determinada por la orientación de los ejes laterales. Cuando el eje de curvatura cóncavo es horizontal (*izquierda*) el espejo invierte las 9:00 y las 3:00 pero no las 6:00 y las 12:00, y se forma una imagen derecha y no trastrocada. Si el espejo se hace girar un ángulo de 45 grados en el sentido de las agujas del reloj (*derecha*), el espejo invierte las direcciones 1:30 y 7:30, pero no las 4:30 y 10:30. La imagen que proporciona el reloj gira 90 grados en el mismo sentido (es decir, un ángulo doble que el girado por el espejo).

en un punto único, sino a lo largo de una línea paralela a la curvatura plana del cilindro cóncavo. Un observador situado más lejos del espejo de lo que se encuentra esta línea focal verá una imagen no trastrocada como resultado de la focalización de frentes de onda invertidos y no simétricos, pero en general la imagen resultante será borrosa y distorsionada. La imagen se hace más nítida a medida que el observador se aleja del espejo, debido a que los frentes de onda se van haciendo más simétricos conforme se alejan de la línea focal. El cilindro cóncavo no es el único que presenta esta propiedad. Cualquier espejo curvo que carezca de simetría rotacional es capaz de focalizar la luz incidente desde una fuente puntual hasta un lugar geométrico no puntual. De aquí que las imágenes reflejadas en espejos tales como el cilindro convexo resulten también afectadas por esta pérdida de nitidez. El astigmatismo ocular está relacionado con este mismo fenómeno.

Para clarificar todas estas cuestiones es conveniente hacer una distinción entre dos tipos de imágenes. Llamaré imagen primaria a la formada por un espejo solo; esta imagen aparece "suspendida" en el espacio y está situada en el punto focal del espejo o (en el caso de los espejos asimétricos) en el lugar geométrico de los puntos focales. Denominaré imagen secundaria a la que forma un espejo en relación con un observador dado; se puede decir que esta imagen se sitúa en el ojo del observador (en la superficie de la retina o en una placa fotográfica). El concepto de imágenes primarias y secundarias nos sirve para explicar la forma en que un espejo del tipo de cilindro cóncavo puede combinar las propiedades de un espejo plano, que refleja imágenes virtuales, y de un espejo cóncavo, que (para observadores y objetos suficientemente alejados del espejo) refleja imágenes reales. ¿Cómo actúa una imagen que parece estar en dos lugares a la vez? Este problema de percepción profunda, conocido como paradoja del paralaje, se resuelve a partir de la idea de que la imagen no está frente al espejo ni detrás de él, sino que se encuentra en el ojo del observador.

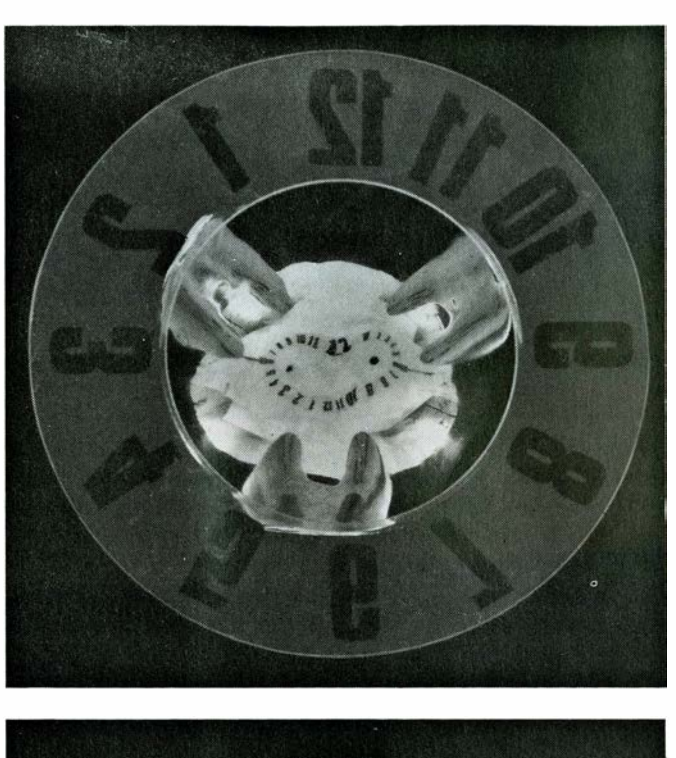
Es importante comprender que las imágenes no trastrocadas que origina un cilindro cóncavo son el resultado de una reflexión simple de la luz. Un par de espejos planos unidos según ángulos rectos puede originar también imágenes no trastrocadas, pero, en este caso, las imágenes se forman como resultado de dos reflexiones independientes y, por tanto, de dos inversiones de la luz independientes entre sí. De hecho, al situar un objeto frente a una combinación de espejos de este tipo, se forman tres imágenes: además de las imágenes derechas y trastrocadas que engendra cada uno de los dos espejos, aparece una tercera imagen no trastrocada, como resultado de una doble reflexión, en el lugar en que se unen los espejos en cuestión.

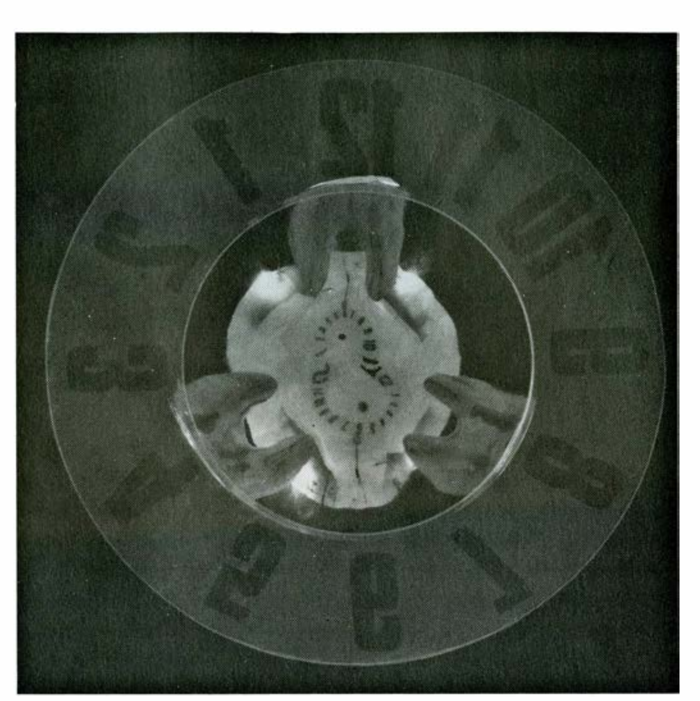
El cilindro cóncavo no es el único espejo curvado básico en el que se forman imágenes no trastrocadas como resultado de una simple reflexión de la luz. Vamos a tomar en consideración el espejo en forma de silla de montar. Este espejo viene a comportarse como un espejo cóncavo a lo largo de uno de los ejes de curvatura principales y como un espejo convexo a lo largo del otro. Al igual que sucedía en el cilindro cóncavo, para un observador o un objeto suficientemente próximo a su superficie, el espejo con forma de silla de montar da lugar a imágenes derechas, trastrocadas y cuyo tamaño resulta reducido a lo largo del eje de curvatura convexa y aumentado a lo largo del eje de curvatura cóncava. Cuando observador y objeto están lejos del espejo, las imágenes que se forman dependen de la orientación del espejo.

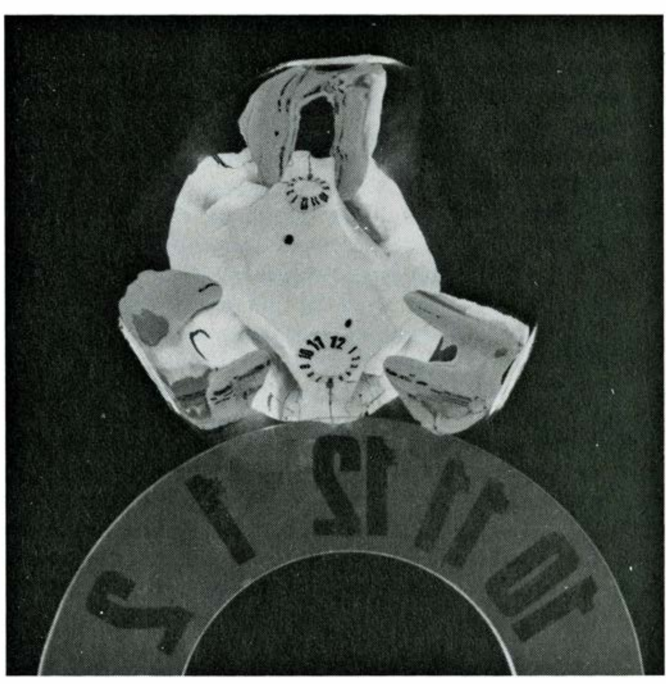
La superficie de un espejo en forma de silla de montar puede dividirse en dos colinas y dos valles, dispuestos de manera que los centros de las colinas están situados sobre el eje de la curvatura cóncava y los centros de los valles sobre el eje de la curvatura convexa.

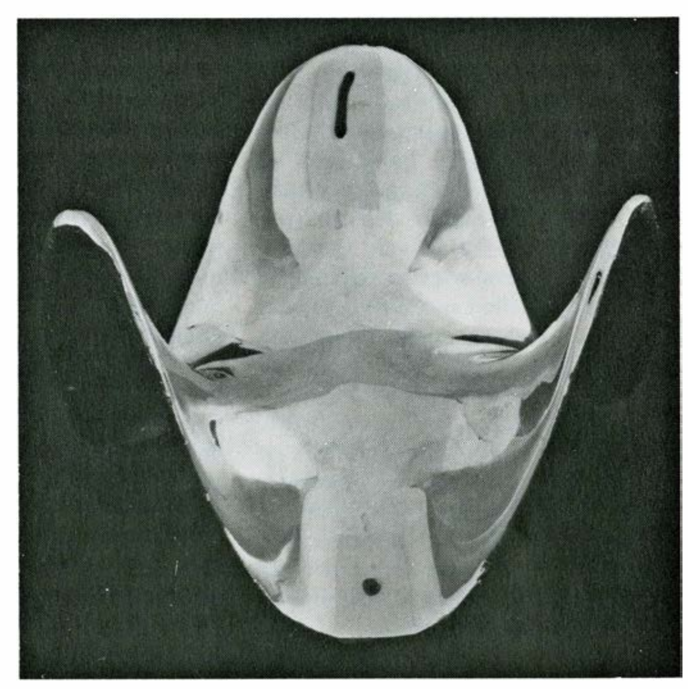
De aquí que las direcciones laterales que atraviesan el centro de una colina se invierten, en tanto que las direcciones laterales que atraviesan el centro de un valle no se invierten. Si el espejo al que nos estamos refiriendo se dispone de forma tal que sus colinas se sitúan a lo largo de una línea horizontal, las 9:00 y las 3:00 serán direcciones de inversión, mientras que las 12:00 y las 6:00 serán direcciones de no inversión. Puesto que las 12:00 y las 6:00 no se invierten, la imagen de un objeto será derecha; dado que la orientación se invierte únicamente a lo largo de una dirección lateral, la imagen no será trastrocada. Por otra parte, en comparación con la imagen que proporciona un espejo plano, el tamaño de esta imagen no trastrocada será menor, en general, para todas las direcciones laterales. (Sin embargo, es posible que la imagen aparezca aumentada a lo largo de la dirección de la curvatura cóncava.)

Al hacer girar un espejo con forma de silla de montar un ángulo de 45







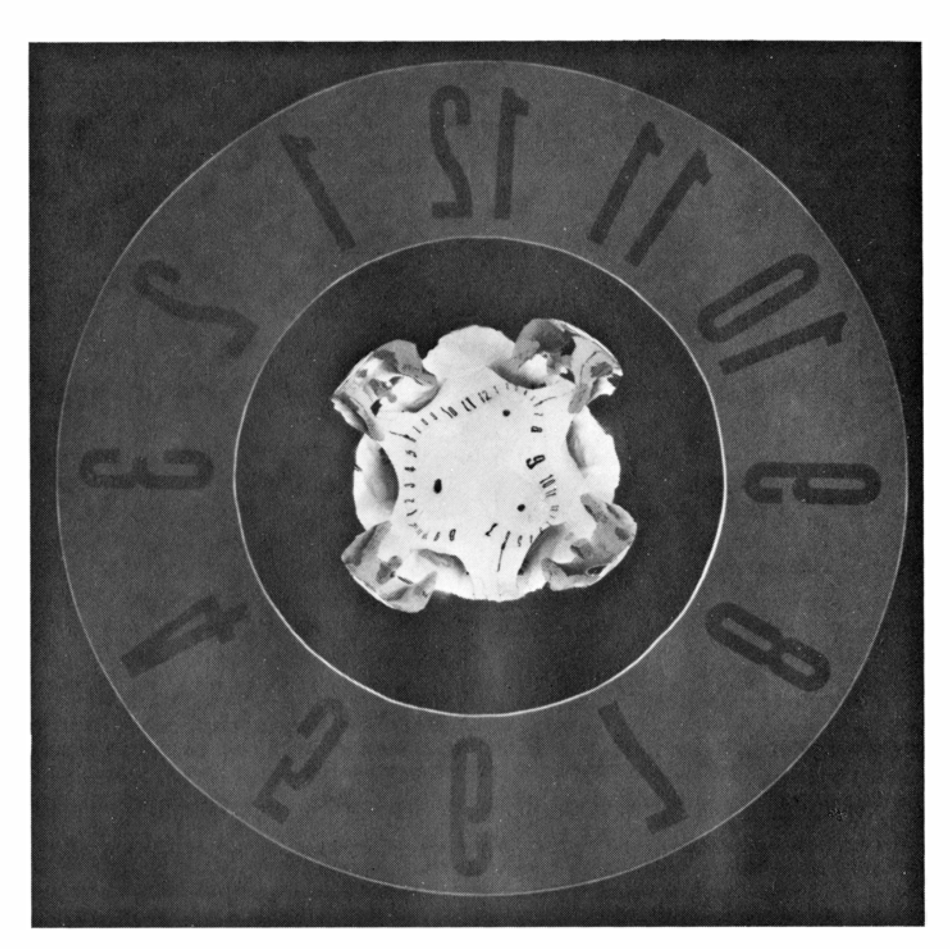


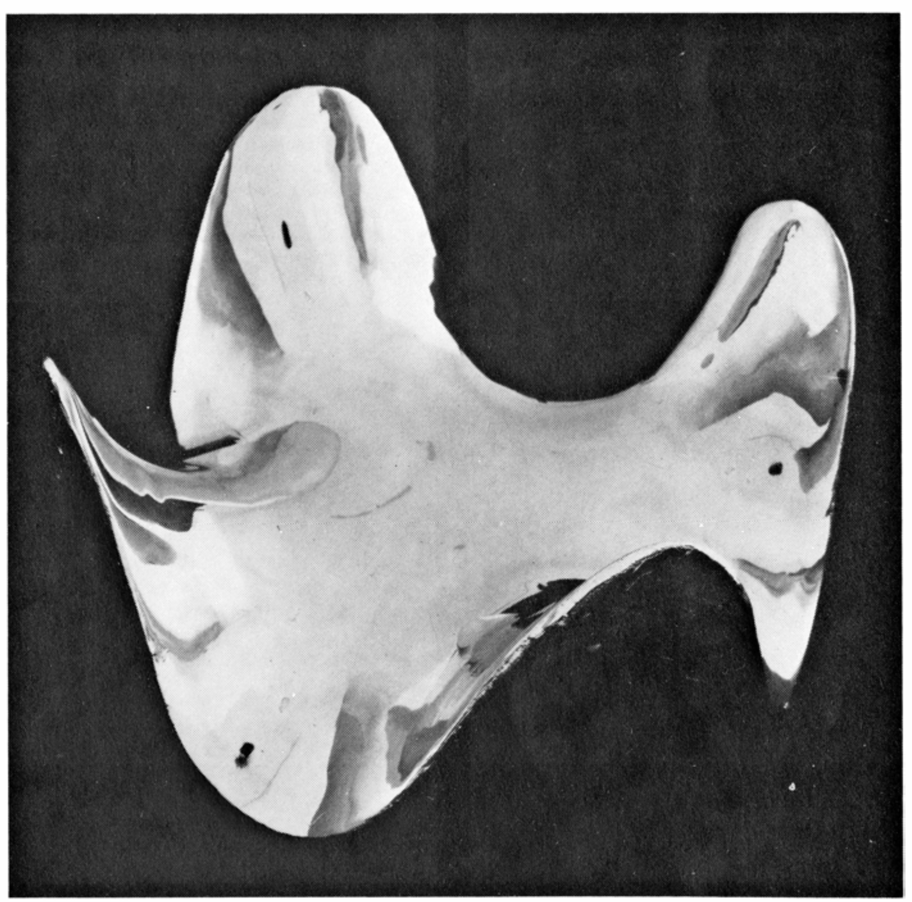
POR ESPEJO MULTILOBULAR, tal como la versión de tres lóbulos que se representa en esta figura, se entiende un espejo compuesto originado por la rotación de una superficie con forma de silla de montar, integrado por dos lóbulos, alrededor de una superficie plana. Tal como sucede con el espejo con forma de silla de montar, el espejo multilobular proporciona imágenes no trastrocadas e invierte la orientación de aquellas direcciones que atraviesan el centro de una colina, pero no la de aquellas que atraviesan el centro de un valle. Cuando el espejo se coloca en la forma que se indica en la figura superior izquierda, las direcciones 2:00, 6:00 y 10:00 se invierten y, en consecuencia, el 8 aparece en la 2:00, el 12 en la 6:00 y el 4 en la 10:00. Por otra parte, las direcciones 12:00, 4:00 y 8:00 no se invierten y, por tanto, aparece otro 12 en la 12:00, otro 4 en la 4:00 y otro 8 en la 8:00. Este espejo forma dos imágenes no trastrocadas de cada número. El resultado global es la esfera de un reloj de 24 horas no trastrocadas. Cuando el espejo trilobular gira 60 grados en el sentido de las agujas del reloj, tal como se representa en la figura superior derecha, las direcciones 12:00, 4:00 y 8:00 se invierten y las direcciones 2:00, 6:00 y 10:00 no se invierten. La imagen de la esfera del reloj de 24 horas aparece girada un ángulo de 90 grados en el sentido de las agujas del reloj. Cuando el reloj objeto se aleja del eje del espejo, la esfera de 24 horas reflejada se divide en dos relojes de 12 horas cada uno, no trastocados.

grados en el sentido en que se mueven las agujas del reloj, los centros de las colinas pasan de las 4:30 a las 10:30, de forma tal que estas direcciones se invierten, en tanto que las direcciones correspondientes a las 1:30 y las 7:30 no se invierten. En este caso, el 12 aparece en las 3:00 y el 3 en las 12:00, el 6 aparece en las 9:00 y el 9 en las 6:00. La imagen todavía es no trastrocada, pero como ha girado 90 grados parece estar en su sitio. Si el espejo gira otros 45 grados, de suerte que los centros de las colinas estén en una línea vertical y los centros de los valles en una línea horizontal, las 12:00 y las 6:00 estarán invertidas, pero las 9:00 y las 3:00 no. En otras palabras, la imagen no trastrocada estará invertida, con su parte superior hacia abajo. En general, si un espejo con forma de silla de montar gira un cierto ángulo, la imagen de un objeto fijo gira en el mismo sentido en una cuantía aproximadamente igual al doble de dicho ángulo.

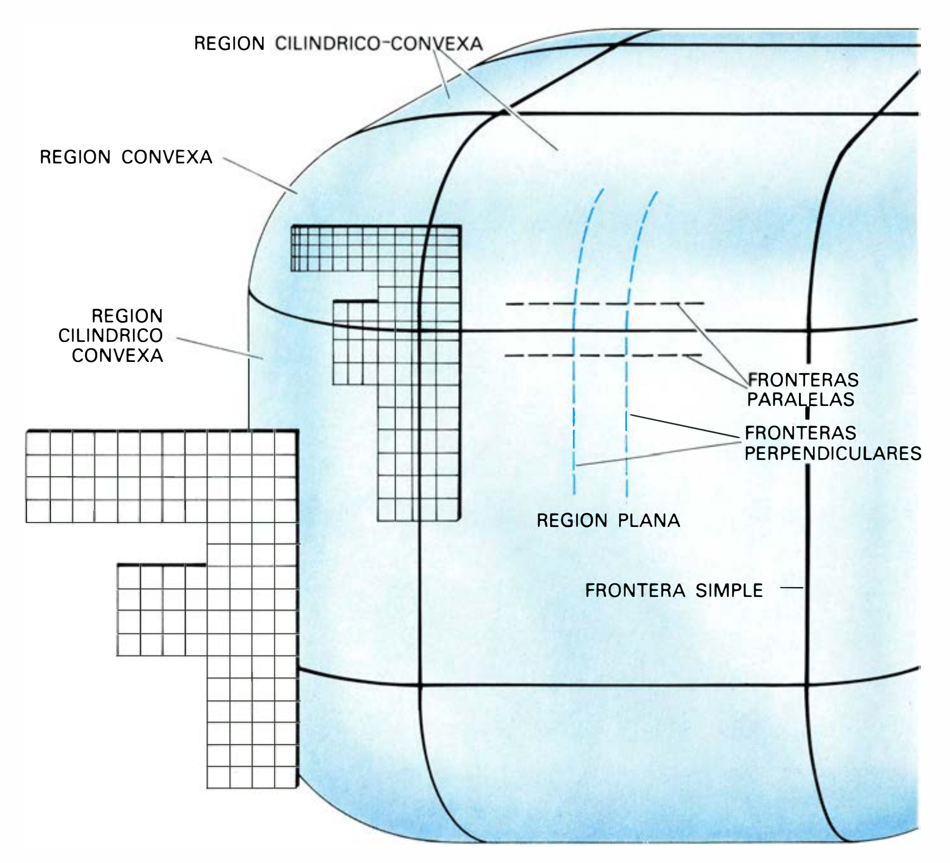
Las propiedades de este tipo de espejos se pueden ilustrar con más claridad si imaginamos la forma en que tales espejos reflejan la esfera de un reloj transparente equipado con una saeta horaria. Cuando la saeta real realiza una vuelta completa de 12 horas alrededor de toda la esfera, la saeta imagen realizará también una rotación completa alrededor de la imagen de la esfera no trastrocada, pero (desde el punto de vista del observador) dicha imagen se moverá en sentido opuesto. Sin embargo, si el propio espejo gira en un sentido contrario al que se mueven las agujas del reloj, de modo que complete una rotación cada 24 horas, el movimiento de la saeta imagen debido a la rotación del espejo y el movimiento de la saeta imagen debido a la rotación de la saeta objeto se compensan mutuamente. De hecho, la imagen de la saeta horaria en movimiento parece estar quieta, en tanto que la imagen de la esfera del reloj, que no se mueve, completa una rotación en sentido contrario al de las agujas del reloj cada 12 horas.

Combinando adecuadamente los espejos de los seis tipos básicos se puede crear una enorme variedad de espejos compuestos con curiosas propiedades reflectoras. Por ejemplo, personalmente he diseñado y patentado una serie de espejos diferentes, empezando por uno con forma de silla de montar que puede engendrar una imagen no trastrocada o varias imágenes de este mismo tipo a partir de un simple objeto individual. Aunque su tamaño es menor que las

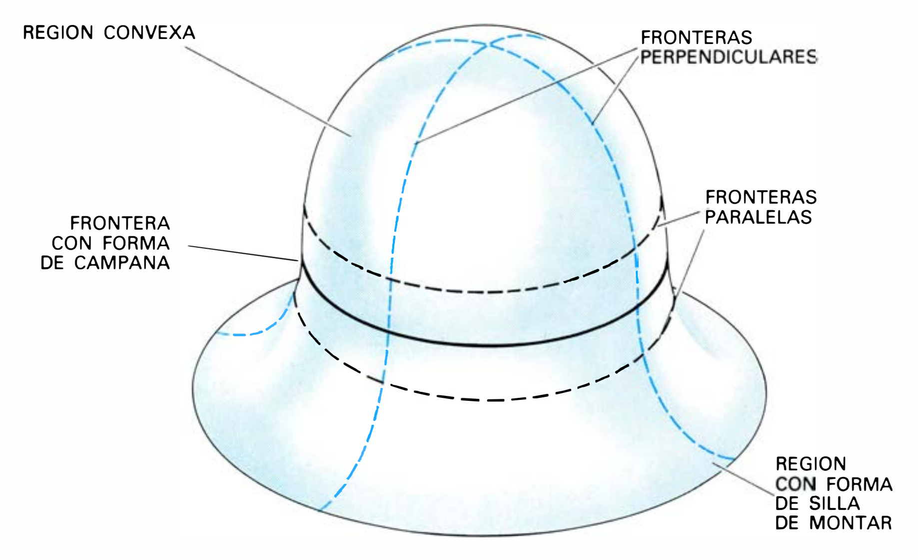




ESPEJO CON CUATRO LOBULOS, representado aquí tal como se ve desde arriba y desde un lado; proporciona tres imágenes no trastrocadas de un objeto dado. En la presente figura el espejo se ha colocado de forma tal que invierta las direcciones 1:30, 4:30, 7:30 y 10:30, pero no las direcciones 3:00, 6:00, 9:00 y 12:00. El resultado a que se llega es una imagen no trastrocada de un reloj con 36 horas.



REGIONES PRINCIPALES DE UN ESPEJO COMPUESTO. Dichas regiones se unen entre sí mediante fronteras suaves (*líneas negras de trazo continuo*). Estas líneas se pueden clasificar de acuerdo con los efectos observados sobre dos conjuntos de curvas trazadas sobre la superficie del espejo. Estas curvas son las fronteras paralelas (*líneas negras de trazo discontinuo*) y las fronteras perpendiculares (*líneas coloreadas de trazo discontinuo*), que relacionan las regiones situadas a ambos lados de la citada frontera. El espejo con forma de caja y aristas redondeadas que se ha representado en esta figura está constituido por tres tipos de regiones: plana, convexa y cilindro-convexa. En todos los casos, al movernos desde uno de los lados de la frontera al otro, las propiedades inversoras de las fronteras paralelas y de las fronteras perpendiculares permanecen inalteradas. Por ejemplo, en la parte superior de la cara anterior, donde una región plana se une con una región cilíndrico-convexa, las fronteras paralelas son planas y, por consiguiente, no inversoras a ambos lados de la frontera; las fronteras perpendiculares cambian de plana a convexa, pero estas dos curvaturas no son inversoras. Una frontera que no cambia las propiedades inversoras de sus paralelas y de sus perpendiculares se conoce con el nombre de frontera simple; un espejo compuesto que sólo contenga fronteras simples refleja una imagen única de un objeto único. Sin embargo, el tamaño de las diferentes partes de la imagen puede variar según las regiones de reflexión.



FRONTERA del tipo de campana (*línea negra continua*). Este límite cambia las propiedades inversoras de las fronteras perpendiculares (*líneas coloreadas discontinuas*), pero no las de las fronteras paralelas (*líneas negras discontinuas*). Una frontera de este tipo une una región convexa con una región con forma de silla de montar. De ello resulta un espejo con forma de campana. Dado que las propiedades inversoras de las fronteras perpendiculares se ven alteradas, en cada par de regiones se forma un par de imágenes.

que corresponderían a un espejo plano, las imágenes que proporciona este espejo poseen una nitidez extraordinaria. Estos espejos compuestos tienen su origen en el espejo con forma de silla de montar y podemos suponer que se generan a partir de este tipo de espejos, girándolos (de manera un tanto compleja) respecto a una cierta región plana de dimensiones muy pequeñas. De este modo, se puede generar una superficie con forma de silla de montar con cualquier número de lóbulos o colinas. A estas superficies reflectoras las denominaré espejos multilobulares.

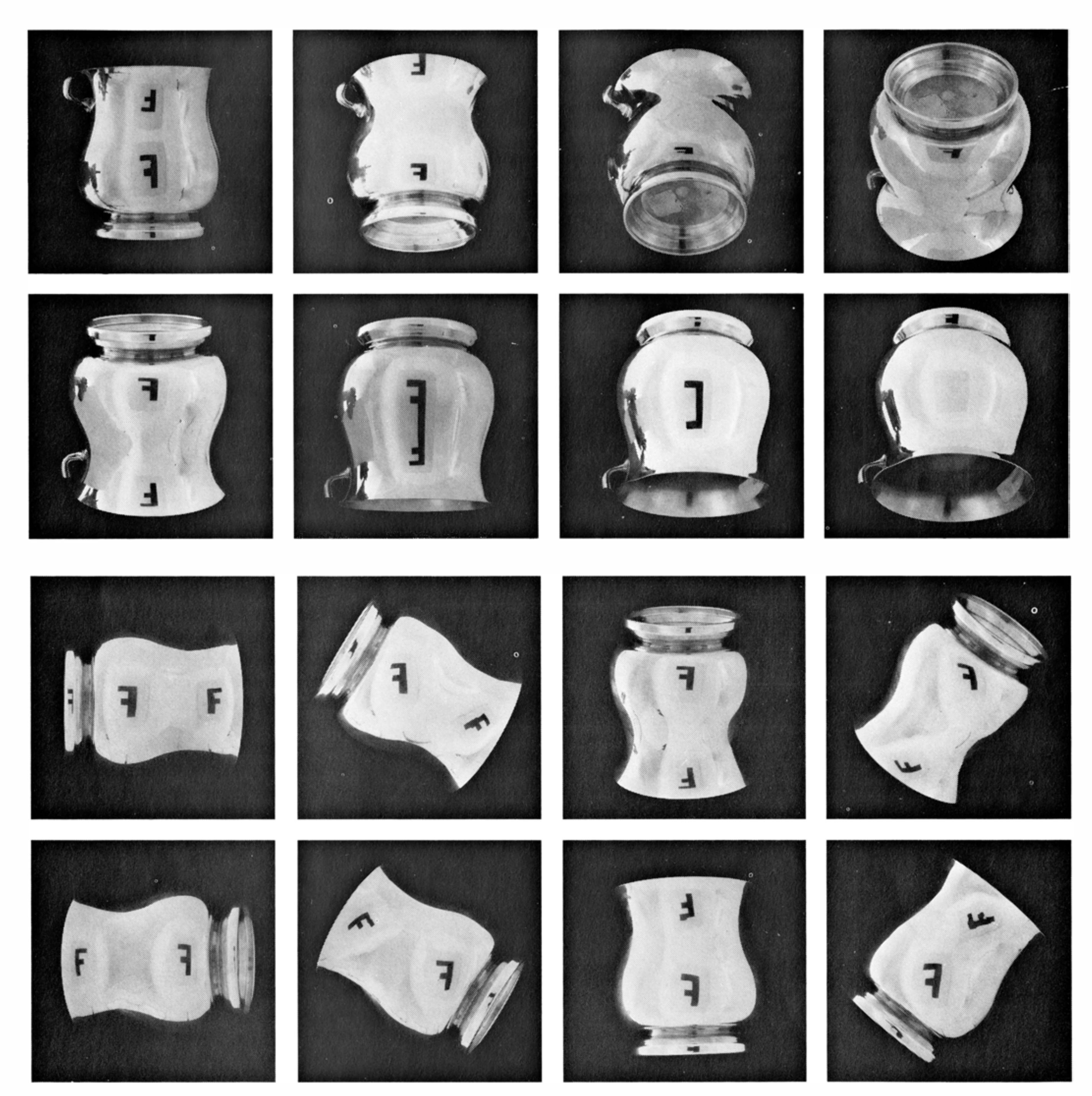
Tal como ya he demostrado en el caso del espejo con forma de silla de montar, o espejo de dos lóbulos, las direcciones laterales que, partiendo de un centro de curvatura, pasan a través del centro de una colina en un espejo multilobular son direcciones de inversión, en tanto que las que atraviesan el centro de un valle son direcciones de no inversión. Cuando un espejo trilobular se coloca con sus colinas según las posiciones 2:00, 6:00 y 10:00, invertirá estas direcciones, pero no sucederá así con las 12:00, 4:00 y 8:00, que son las direcciones que corresponden a los valles. Debido a las inversiones, aparece un 8 en las 2:00, un 12 en las 6:00 y un 4 en las 10:00. Sin embargo, debido a las reflexiones que no producen inversión, aparece otro 12 en las 12:00, otro 4 en las 4:00 y otro 8 en las 8:00. De hecho, si la superficie del espejo es continua y suave, surgen dos imágenes para cada una de las 12 cifras que hay grabadas en la esfera del reloj que constituye nuestro objeto. Dado que cada uno de los dos conjuntos de cifras imágenes aparece en el mismo orden en que se mueven las saetas de un reloj, la imagen de la esfera de un reloj de 12 horas es una esfera de un reloj de 24 horas.

Si el espejo trilobular gira 60 grados, invierte las direcciones 12:00, 4:00 y 8:00, pero no las 2:00, 6:00 y 10:00. En este caso, la imagen todavía es la de un reloj de 24 horas, pero tiene un 12 en las 3:00 y otro en las 9:00, un 10 en las 4:00 y otro en las 10:00, y así sucesivamente. En general, si el espejo trilobular gira en una cuantía determinada, la imagen correspondiente a un objeto fijo gira en la misma dirección en una cuantía que resulta ser una vez y media mayor que la anterior.

La saeta horaria de un reloj transparente pone de manifiesto un aspecto todavía más sorprendente de este mismo fenómeno. Cuando dicha saeta gira a lo largo de las 12 horas de la esfera

del reloj, se observan dos saetas imágenes moviéndose a lo largo de una esfera no trastrocada de 24 horas. Cada saeta recorre 180 grados cada 12 horas. Tras haber transcurrido 12 horas, las dos saetas imágenes han intercambiado sus posiciones y únicamente después de que la saeta horaria real haya completado dos rotaciones de 12 horas, cada una de las saetas imágenes habrá vuelto a su posición original. Resulta también interesante señalar que cuando un reloj con la esfera transparente se mueve hasta uno de los lados del espejo de tres lóbulos de forma tal que el centro de la esfera no esté ya alineado con el eje del espejo, la esfera imagen no trastrocada de 24 horas se divide en dos esferas no trastrocadas de 12 horas cada una.

Un espejo con forma de silla de montar con n lóbulos (donde n es igual o mayor que dos) puede crear $n - 1$ imágenes no trastrocadas de un objeto individual. En consecuencia, un espejo con cuatro lóbulos puede dar lugar a tres imágenes no trastrocadas y, en particular, la imagen de una esfera de reloj de 12 horas será una esfera con 36 horas. Por otra parte, cuando un espejo con n lóbulos gira en una cuantía determinada, la imagen de un objeto fijo girará, en el mismo sentido, $n/(n - 1)$ veces dicha cantidad. Por el contrario, si el objeto que se refleja en un espejo fijo con n lóbulos se hace



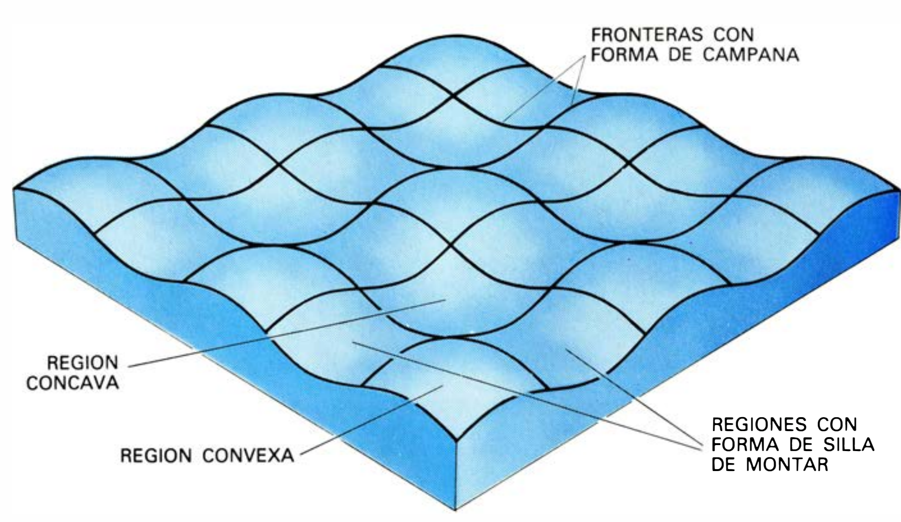
A AMBOS LADOS DE UNA FRONTERA del tipo de campana aparecen sendas imágenes enantiomórficas (es decir, imágenes con lateralidad opuesta). Un espejo de esta clase proporciona una imagen derecha y trastrocada sobre su región convexa y una imagen invertida y no trastrocada sobre su región con forma de silla de montar. La secuencia de fotografías que se reproducen en la parte superior de esta figura pone de manifiesto lo que sucede con las imágenes que refleja este espejo al girarlo alrededor de un eje lateral: en un principio, las dos imágenes se aniquilan entre sí y, más tarde, reaparecen de nuevo. La secuencia inferior muestra lo que sucede cuando se gira el espejo en el mismo sentido en que se mueven las agujas de un reloj: la imagen trastrocada se mantiene derecha, en tanto que la no trastrocada sufre dos rotaciones completas en el curso de una rotación única del espejo.

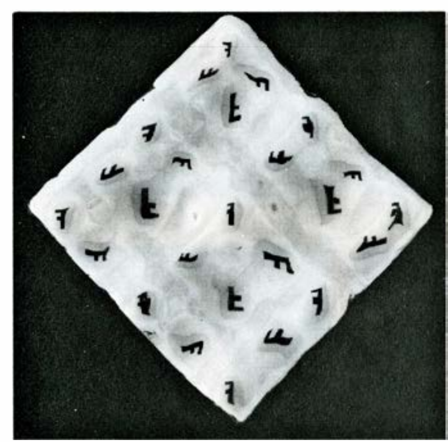
girar en una cierta cuantía, su imagen girará, en sentido opuesto, $1/(n-1)$ veces dicha cantidad.

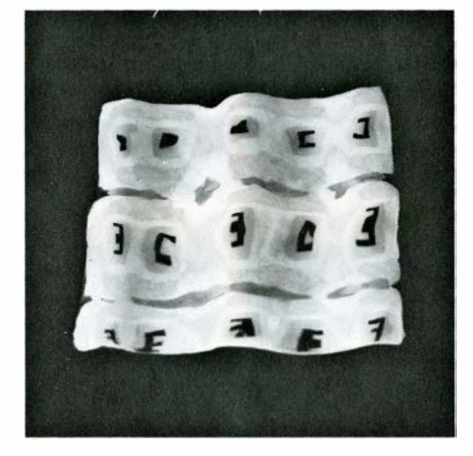
Los espejos con forma de silla de montar son simplemente uno de los muchos tipos de espejos compuestos que son capaces de formar más de una imagen a partir de un simple objeto. De hecho, dado el número ilimitado de combinaciones que se pueden construir con los diferentes tipos de espejos básicos que existen, la clasificación de los espejos compuestos que producen imágenes múltiples se presenta como una tarea casi imposible. Sin embargo, el problema no es tan difícil como podría parecer en un principio. La clave de su solución consiste en tomar en consideración no sólo las regiones básicas que integran un espejo compuesto, sino también las zonas que separan a una región de otra. (Se supone que la "unión" entre dos regiones básicas de un espejo compuesto es suave, es decir, sin discontinuidades.) Las fronteras de un espejo compuesto se pueden clasificar en tres categorías principales. Identificando la categoría a que pertenece una determinada frontera, podremos predecir las imágenes que se formarán a ambos lados de la misma.

La mejor forma de describir las fronteras de un espejo compuesto consiste en considerar su relación con dos conjuntos de curvas que se sitúan sobre la superficie del espejo: las curvas que van paralelamente a la frontera y las que van perpendicularmente a ella. Como cualquier otra curva que podamos imaginar sobre la superficie del espejo, las fronteras paralelas y las fronteras perpendiculares pueden ser inversoras (si su curvatura es cóncava) o no inversoras (si su curvatura es convexa o plana). En consecuencia, la frontera entre dos regiones básicas se puede clasificar de acuerdo con el cambio que experimentan las propiedades inversoras de las correspondientes líneas paralelas y perpendiculares al pasar de un lado de la frontera al otro.

Por ejemplo, me referiré a una frontera simple si, al pasar de un lado al otro, ni las paralelas ni las perpendiculares se transforman de curvas inversoras en curvas no inversoras (o viceversa). Los espejos compuestos cuyas regiones están separadas por fronteras simples forman imágenes únicas al poner frente a ellas un objeto dado. Las diversas partes de la imagen pueden presentar tamaños distintos en las diferentes regiones del espejo, en razón de la curvatura que presentan tales regiones; empero, cada una de dichas partes está relacionada con las demás en la misma forma en que están relacionadas en el objeto real.





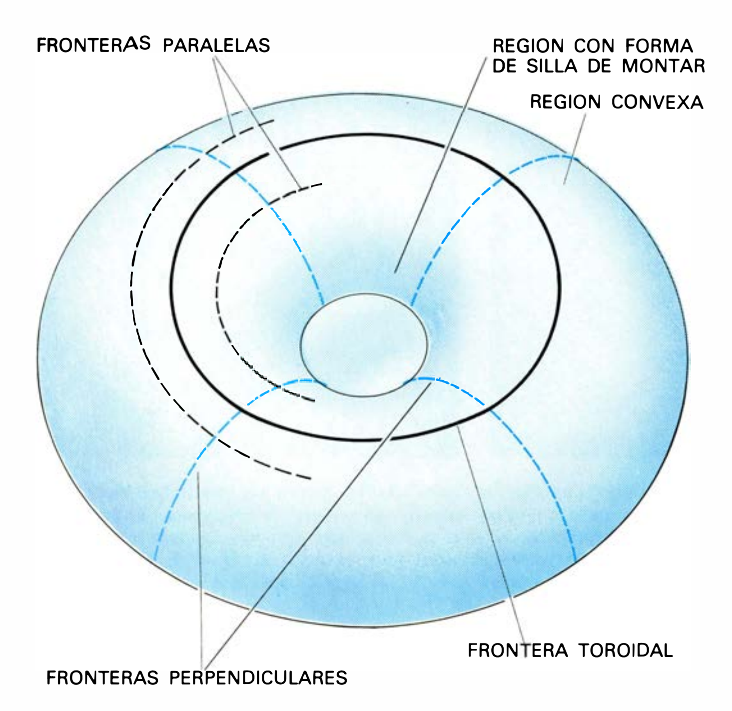


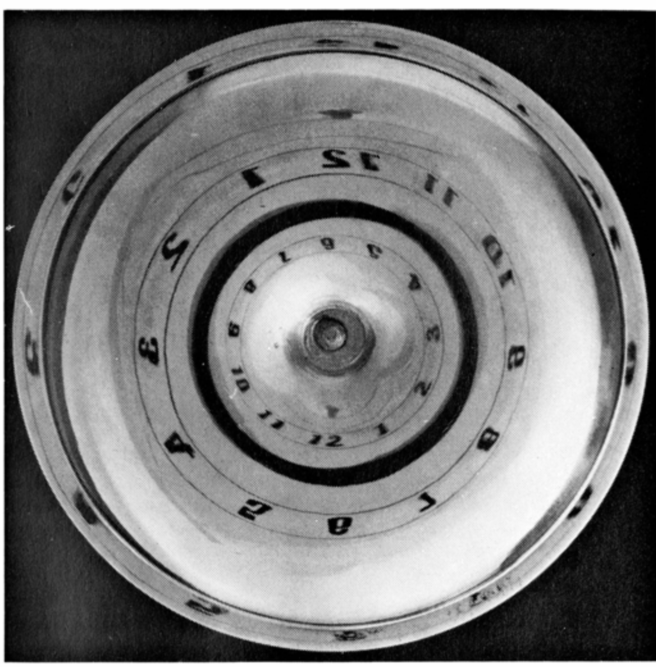




UN ESPEJO DE MATRIZ consiste en cuatro o más regiones básicas unidas por fronteras con forma de campana. El ejemplo que se reproduce en esta figura está constituido por regiones convexas, cóncavas y con forma de silla de montar, repetidas en una secuencia de dos por dos (*como se señala en el dibujo*). Dado que existen fronteras del tipo de campana al menos sobre dos de los lados de cada región básica, se puede formar un grupo de cuatro imágenes en cada celda de la matriz dos por dos. Cuando el espejo se inclina lateralmente se crean y aniquilan una serie de pares dobles de imágenes enantiomórficas.

Una frontera cuyas perpendiculares cambian su carácter de inversor a no inversor, pero cuyas paralelas no cambian dicho carácter, recibe la denominación de frontera de campana. Como cabría esperar, un espejo con forma de campana consta de una región convexa unida a una región con forma de silla de montar y constituye un ejemplo de espejo compuesto con una frontera de este tipo. Debido al hecho de que las propiedades inversoras de las perpendiculares cambian en la frontera, por cada par de regiones se crea un par de imágenes. Estas imágenes son enantiomórficas (tienen las derechas y las izquierdas opuestas) y están situadas





FRONTERA TOROIDAL. Cambia las propiedades inversoras de las fronteras paralelas, pero no las de las fronteras perpendiculares (*líneas coloreadas de trazo discontinuo*). En el espejo con forma toroidal que se representa en el diagrama de la izquierda, la frontera toroidal separa la región convexa, situada en el exterior, de la región con forma de silla de montar, situada en el interior. Estas dos regiones crean pares de imágenes con lateralidad opuesta y separadas entre sí 180 grados. En la fotografía de la derecha se puede observar que la esfera de un reloj se refleja formando dos imágenes concéntricas. Una de estas imágenes es trastrocada y la otra no lo es; los números que aparecen en ambas imágenes se presentan en un orden contrario al de las agujas del reloj. El reloj que reproduce la imagen interior parece haber sido puesto al revés, con la parte de dentro hacia fuera.

simétricamente a ambos lados de la frontera de la campana.

Cuando un espejo con forma de campana (o cualquier otro espejo compuesto capaz de engendrar pares de imágenes enantiomórficas) va girando alrededor de un eje lateral paralelo a la frontera de la campana, las imágenes a que da lugar se pueden combinar y aniquilarse mutuamente. Al girar el espejo en sentido contrario, se reproduce el par de imágenes anterior, en apariencia surgiendo de la nada. Sin embargo, cuando un espejo del tipo de campana gira alrededor de su eje anteroposterior, las imágenes enantiomórficas se comportan de modo muy distinto.

Por una parte, dado que las imágenes trastrocadas no se ven afectadas por los cambios en la orientación del espejo, la imagen trastrocada reflejada en la región convexa mantiene la misma orientación cuando se produce una rotación completa del espejo. Por otra parte, puesto que la orientación de las imágenes no trastrocadas depende de la orientación del espejo, la imagen no trastrocada que aparece en la región con forma de silla de montar sufre dos rotaciones completas por cada rotación del espejo. De este modo, las imágenes trastrocadas y no trastrocadas ofrecen siempre lados correspondientes con la frontera de la campana y entre sí.

Un conjunto de cuatro o más regiones diferentes, unidas por fronteras con forma de campana, constituye lo que yo denomino un espejo de matriz. El espejo más simple de este tipo es el que corresponde a una matriz de dos por dos: consiste en dos regiones con forma de silla de montar unidas por un punto y separadas por una región convexa en uno de los lados y por una región cóncava en el otro. Continuando en esta línea, se puede construir un espejo de matriz con cualquier número de filas y columnas. Por ejemplo, una matriz de cinco por cinco es capaz de reflejar hasta 25 imágenes de un objeto dado. Las cuatro fronteras con forma de campana que separan las diferentes regiones de un espejo de matriz dos por dos se reúnen en un punto; al parecer, en este punto coinciden solamente una región convexa y una región cóncava. Este hecho explicaría por qué las fronteras no cambian las propiedades de inversión de las líneas paralelas y perpendiculares a dichas fronteras.

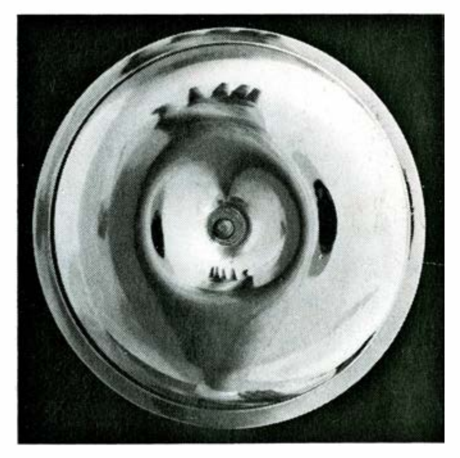
Al igual que sucede en los espejos con forma de campana, los espejos de matriz pueden engendrar pares de imágenes enantiomórficas. Inclinando oblicuamente un espejo de matriz, se pueden formar pares dobles; estos pares se aniquilan o crean simultáneamente. Sin embargo, ningún espejo compuesto puede dar lugar a más de cuatro imágenes de un objeto único en un mismo punto.

La región con forma de silla de montar y la región plana de un espejo multilobular se unen también por fronteras con forma de campana que producen la aparición de algunas imágenes interesantes. Por ejemplo, aunque un espejo trilobular tan sólo refleja normalmente dos imágenes no trastrocadas de un objeto, si un observador mantiene un espejo con una colina en la dirección de las 12:00 y empieza a girar la parte superior del espejo hacia él, las dos imágenes no trastrocadas se mueven hacia la región plana del espejo. Cuando las dos imágenes se aproximan a la región plana, una de ellas se divide en tres nuevas imágenes, dos de ellas no trastrocadas y una trastrocada. Si el espejo gira de nuevo, la imagen trastrocada puede aniquilar cualquiera de las tres imágenes no trastrocadas y, de este modo, tan sólo subsisten dos imágenes no trastrocadas.

El tercer tipo de frontera que existe en los espejos compuestos es aquella que cambia las propiedades de inversión de las líneas paralelas a ella, pero no el de las líneas perpendiculares. A esta clase de fronteras yo las denomino fronteras toroidales, debido a que separan las dos regiones que constituyen un espejo con forma de toroide, es decir, una región exterior convexa y una región interior con







FORMA EN QUE SE COMPORTAN LAS IMAGENES que proporciona un espejo toroidal. Dicho patrón se pone de manifiesto en esta figura con la ayuda de una mano reflejada. Al mover la mano sobre el espejo, la imagen no trastrocada que aparece en la región con forma de silla de montar sufre un cambio complejo con la imagen trastrocada que aparece en la región convexa. En un espejo toroidal, el cambio de una imagen por otra difiere bastante en relación con la creación y aniquilación de pares de imágenes en un espejo con forma de campana: en el caso del toroide, las imágenes trastrocadas y no trastrocadas de un simple objeto no se encuentran adyacentes sino que están desplazadas 180 grados entre sí. (Fotografías cedidas por Ralph Morse.)

forma de silla de montar. En realidad, en un espejo toroidal existen dos fronteras toroidales. Una de estas fronteras une la región convexa y la región con forma de silla de montar alrededor de la parte inferior del toroide; la otra frontera une las dos regiones citadas alrededor de la parte superior del toroide. Sin embargo, en relación con el presente análisis, bastará con que consideremos el toroide solamente desde el punto de vista superior, de suerte que las dos regiones básicas formen un par de anillos concéntricos, con la región dél tipo de silla de montar en el interior, la región convexa en el exterior y una única frontera toroidal separando ambas regiones.

Observada en un sentido axial, esta sección de un espejo toroidal crea imágenes dobles de un objeto dado: una imagen trastrocada sobre la región exterior convexa y una imagen no trastrocada sobre la región interna con forma de silla de montar. Aunque su lateralidad es opuesta, estas dos imágenes difieren de los pares enantiomórficos que reflejan los espejos de matriz o con forma de campana en el hecho de que dichas imágenes no están colocadas simétricamente respecto a la frontera que existe entre ellas. Por el contrario, las dos imágenes toroidales aparecen separadas 180 grados sobre sus anillos respectivos. Por ejemplo, cuando el objeto es la esfera de un reloj, el espejo toroidal crea dos imágenes concéntricas, una trastrocada y orientada en un sentido contrario al de las agujas del reloj en la región convexa y la otra no trastocada y orientada también en un sentido contrario al de las agujas del reloj en la región con forma de silla de montar. De aquí que si la cifra 12 de la imagen exterior aparece en las 12:00, el 12 de la imagen interior aparecerá en las 6:00. De hecho, la imagen no trastrocada del reloj da la vuelta de dentro a fuera; los números no trastrocados se presentan en un orden contrario al de las agujas del reloj y su parte superior se sitúa hacia dentro. En un espejo toroidal, la relación entre la imagen trastrocada y la no trastrocada es verdaderamente singular. Por ejemplo, el 12 trastrocado aparece próximo al 6 no trastrocado.

Si los objetos se mueven sobre un espejo toroidal, sus imágenes experimentan un movimiento peculiar en las proximidades de la frontera del toroide. En vez de aniquilarse mutuamente, tal como sucede con las imágenes enantiomórficas que aparecen en un espejo con forma de campana o un espejo de matriz, estas imágenes se combinan entre sí, dando lugar a una imagen extraordinariamente distorsionada en la frontera del toroide. La imagen trastrocada y la no trastrocada de un simple objeto engendran una imagen con forma de anillo; dicho anillo está constituido por cuatro imágenes (dos trastrocadas y otras dos no trastrocadas). Las nuevas imágenes surgen de la imagen original no trastrocada de forma muy similar a como lo hacen las imágenes de los espejos multilobulares.

Las peculiaridades del espejo toroidal me permiten introducir un tema con el que concluiré el presente estudio de las imágenes que forman los espejos curvados. A la vista de la anterior exposición acerca de los espejos compuestos, resulta evidente que las imágenes trastrocadas y las no trastrocadas difieren en algo más que en su lateralidad. Por ejemplo, todas las imágenes no trastrocadas pueden girar respecto a los objetos fijos, en tanto que las correspondientes imágenes trastrocadas se ven obligadas a permanecer o derechas o invertidas. Más aún, las imágenes trastrocadas o no trastrocadas que producen los espejos con forma de campana o del tipo de matriz se ven afectadas, paritariamente, por las creaciones y aniquilaciones de pares o dobles pares de imágenes; en el caso de espejos multilobulares y toroidales, una imagen no trastrocada puede dividirse en tres nuevas imágenes (dos no trastrocadas y una trastrocada), en tanto que su correspondiente imagen trastrocada permanece intacta.

Como es bien sabido, las imágenes trastrocadas y las no trastrocadas pueden crearse o destruirse. Ahora bien, unas y otras aparecen y desaparecen siempre en igual número, incluso en el caso de los espejos multilobulares y toroidales. Para un determinado espejo compuesto, la diferencia entre el número de imágenes trastrocadas y el número de imágenes no trastrocadas es un "número mágico" cuyo valor permanece constante, independientemente del número total de imágenes que existan en un conjunto determinado. Por ejemplo, el espejo de matriz cinco por cinco puede dar lugar a 13 imágenes trastrocadas y 12 imágenes no trastrocadas; su número mágico será, pues, 13 − 12, es decir, uno. Si el espejo se inclina de forma tal que se eliminen entre sí cuatro filas con cinco imágenes cada una, subsisten cinco imágenes individuales; tres de estas imágenes son trastrocadas y las otras dos no trastrocadas y, en consecuencia, el correspondiente número mágico sigue valiendo la unidad.

Juegos matemáticos

Las regularidades observadas en los números primos son reflejo de la ley fuerte de los pequeños números

Martin Gardner

Ensalcemos los números primos en buenhora,
junto a nuestros padres, que nos engendraron;
pues su gloria, su don, su fuerza peculiar
es carecer de divisores, no tener antepasados.
Entre las generaciones de productos,
ellos son adanes.
–HELEN SPALDING

"La ley fuerte de los pequeños números": he aquí el provocativo título de un artículo inédito de Richard Kenneth Guy, matemático de la Universidad de Calgary. Durante muchos años, Guy ha sido el redactor de la sección "Research Problems" (problemas para investigación) de *The American Mathematical Monthly*. Es autor de numerosos artículos especializados, y junto con John Horton Conway y Elwyn R. Berlekamp, coautor de *Winning Ways,* obra colosal dedicada a nuevos problemas de matemática recreativa, que tras larga espera confía que sea publicada este año por Academic Press. Casi todo cuanto sigue ha sido tomado del artículo de Guy.

"Consideramos que la matemática es una ciencia deductiva y exacta", comienza diciéndonos Guy. "Pero en el campo de la investigación y el descubrimiento esta imagen no es en modo alguno válida. En investigación, dos aspectos de máxima importancia estriban en saber formular las preguntas oportunas y en saber reconocer pautas y regularidades."

Por desgracia, no existe ningún procedimiento para generar sistemáticamente preguntas atinadas, y tampoco hay forma de saber si al reconocer una determinada pauta nos veremos abocados a un teorema nuevo e importante, o si la pauta observada se quedará en mera coincidencia. En estos aspectos, los matemáticos se encuentran en situación curiosamente análoga a la de los demás científicos. Ambos se plantean cuestiones, realizan experimentos y analizan los resultados. Hallada una regularidad, ¿volveremos a encontrarla al efectuar nuevas observaciones correspondientes a distintos parámetros –lo que consentiría enunciar una ley general– o se presentarán contraejemplos que contradigan la hipótesis? Es cierto que los matemáticos pueden hacer algo imposible en otras ciencias: demostrar teoremas en un sistema formal. Empero, en tanto no se descubra la demostración, el matemático ha de apoyarse en la inducción empírica, falible, de forma muy parecida al proceder de la ciencia experimental. Y así sucede sobre todo en aquellos problemas combinatorios que impliquen sucesiones infinitas de números.

Es posible que observando situaciones relativas a números pequeños encontremos regularidades que nos hagan pensar en un teorema general. Este proceso de inducción es lo que Guy llama ley fuerte de los pequeños números. Hay ocasiones donde la ley funciona; en otras no. Cuando la pauta detectada se queda en una serie de coincidencias, cosa frecuente, los matemáticos pueden despilfarrar gran cantidad de tiempo en pos de un teorema falso. La ley puede despistarnos también por causa contraria: quizás unos cuantos contraejemplos sean razón para que un matemático desista prematuramente en la búsqueda de un teorema que sería cierto, aunque un poco más complicado de lo que se esperaba.

Los ordenadores electrónicos nos prestan hoy gran ayuda, porque con frecuencia pueden explorar rápidamente hasta números lo suficientemente grandes como para invalidar la conjetura o para aumentar grandemente su probabilidad de ser cierta. A pesar de ello, en muchos problemas de carácter combinatorio los números crecen a ritmo tan fantástico que el ordenador tan sólo permite examinar unos cuantos casos más de los factibles a mano, quedando el matemático enfrentado a un problema extraordinariamente refractario al análisis.

Podrían llenarse muchos volúmenes con ejemplos de cómo la ley fuerte de los pequeños números, con tanta frecuencia, ha conducido hasta importantes teoremas que encandilaron a los investigadores, lanzándolos a la búsqueda de teoremas inexistentes, o los ha despistado, sugiriendo que no hay ningún teorema, cuando sí lo hay, o ha inducido a pensar en teoremas que se podrían creer al alcance de la mano, pero que resisten incólumes todos los intentos encaminados a demostrarlos. En el revoltillo de ejemplos que siguen centraremos la atención en los números enteros positivos primos.

Se llaman números primos a los enteros naturales mayores que 1 y que tan sólo son divisibles por sí mismos y por la unidad: 2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31,... Todos son impares, excepto el 2, reputado de ser el más "primo" de los primos. Ninguna fórmula polinómica puede generar todos los números primos; más aún, se ha demostrado que no existe ningún polinomio que tome exclusivamente valores primos. Como dice la segunda estrofa del poema de Helen Spalding:

Dispersos entre los ordinales,
¿quién su llegada sabría predecir?
Siempre imprevista:
no ocupan plazas reservadas.
Y al pasar revista
en la procesión de cardinales,
se alzan hieráticos pontífices
uno a uno inescrutables,
uno a uno electos por sí mismos.

Euclides demostró que la sucesión de números primos no tiene fin; mas conforme se va avanzando en ella, también van creciendo los intervalos que los separan. Salvo el 6, todo número natural menor que 10 es potencia de un número primo; más de la

$2 + 1 = 3$	$5 - 2 = 3$
$(2\times3) + 1 = 7$	$11 - 6 = 5$
$(2\times3\times5) + 1 = 31$	$37 - 30 = 7$
$(2\times3\times5\times7) + 1 = 211$	$223 - 210 = 13$
$(2\times3\times5\times7\times11) + 1 = 2311$	$2333 - 2310 = 23$
$(2\times3\times\ldots\times13) + 1 = 30031$	$30047 - 30030 = 17$
$(2\times3\times\ldots\times17) + 1 = 510511$	$510529 - 510510 = 19$
$(2\times3\times\ldots\times19) + 1 = 9699691$	$9699713 - 9699690 = 23$

Conjetura de Fortune

tercera parte de los números menores que 100 son potencias de números primos. Empero, a la vista de estos pequeños números sería ridículo concluir que la densidad de las potencias de los números primos tiene cota inferior. Por el contrario, van enrareciéndose tan rápidamente al avanzar en la sucesión de los números naturales que su densidad desciende por bajo de cualquier valor positivo dado de antemano.

En el principio, donde el caos finaliza
y todo se reduce a cero,
llenan el paisaje, como los árboles el bosque.
Pero la media distancia ya los enrarece,
y a lo lejos, hacia el infinito,
son tan escasos como erráticos cometas.

Los números primos presentan enjundiosos ejemplos de pautas muy notables, totalmente accidentales, que no conducen a parte ninguna. Tomemos la siguiente sucesión de números primos: 7, 37, 337, 3337, 33337, 333337,... Nos sentimos inclinados a pensar que esta misma ley de formación continuará generando números primos; mas ya en el caso siguiente falla la propiedad: 3333337 es compuesto (no primo), siendo expresable como producto de los factores primos 7 × 31 × 15361. La verdad es que en todos los casos de pautas de este tipo podemos apostar con tranquilidad a que la regla no producirá indefinidamente números primos.

Hace algunos años, Reo F. Fortune, antropólogo de la Universidad de Cambridge (estuvo casado con la recientemente fallecida Margaret Mead), hizo notar una curiosa pauta observable en números primos no muy grandes. Empezando por 2, tomemos el producto de un conjunto de primos consecutivos. Se suma 1 al resultado. Se busca el número primo inmediatamente posterior, y se le resta el producto de los primos consecutivos. ¿Será siempre número primo esta diferencia? La tabla de la página adyacente muestra cómo se aplica el procedimiento en los ocho primeros casos, y da los ocho primeros "primos afortunados" generados por este método.

Fortune ha conjeturado que su método genera siempre números primos. Casi todos los especialistas en teoría de números admiten que así debe suceder, pero no se ha podido dar ninguna demostración, y, según dice Guy, no parece haber mucha esperanza de que se logre alguna en plazo previsible. Quizás algún lector de la sección pueda "cocer" (demostrar la falsedad de) la conjetura, atinando con lo que podríamos llamar un "cocinero de fortuna". Observemos que, en la tabla, los cinco primeros números del segundo miembro de la ecuación de la izquierda son primos. ¿Ocurre siempre así? No, pues falla en los tres números siguientes. En un artículo titulado "On the Primality of $k! + 1$ y $2^* 3^* 5^* \dots ^* p + 1$" Mark Templer ha demostrado (*Mathematics of Computation,* vol. 34, n.º 149, págs. 303-304; enero de 1980) que sumando una unidad al producto de los números primos consecutivos hasta p resulta un número primo para los cinco primeros primos, así como para $p = 37$, $p = 379$, $p = 1019$ y $p = 1021$, y que esta cantidad ya no vuelve a ser número primo para valores de p menores que 1032.

Otra extraña hipótesis, tampoco demostrada todavía, es la llamada conjetura de Gilbreath, en honor de Norman L. Gilbreath, matemático norteamericano y mago aficionado, quien la dio a conocer en 1958. Escribamos en hilera la sucesión de los números primos, y, bajo ellos, anotemos la lista de diferencias entre primos consecutivos. Debajo de esta segunda hilera dispongamos los valores absolutos de las diferencias, prosiguiendo de esta forma tanto como se desee. La ilustración de esta página muestra una tabla de nueve hileras de diferencias correspondientes a los primeros 24 números primos. Observemos que cada hilera comienza por 1. ¿Les ocurriría así a todas? Gilbreath presume que sí, lo que ha sido comprobado por Ray B. Killgrove y Ken E. Ralston hasta el número primo de lugar 63.419 (*Mathematical Tables and Other Aids to Computation,* vol. 13, n.º 66, págs. 121-122; abril de 1959).

"No parece verosímil que podamos ver una demostración de la conjetura de Gilbreath en fecha cercana, a pesar de que probablemente sea cierta", nos previene Guy, quien añade que de ser verdadera es posible que sea independiente de que los números de partida sean primos. Hallard Croft ha sugerido que la conjetura pudiera ser válida para cualquier sucesión que comience por 2, seguido de números impares que vayan creciendo a ritmo "razonable" separados por lagunas de tamaño también "razonable". De ser así, la hipótesis de Gilbreath pudiera no ser tan misteriosa como parece a primera vista, pese a que pueda resultar enormemente difícil de demostrar.

Una de las más notables entre todas las conjeturas sobre números primos que se encuentran pendientes de solución es la que afirma que existe un número infinito de primos gemelos, es decir, de números primos que se diferencian en dos unidades. Los ejemplos de más bajo valor son 3 y 5, 5 y 7, 11 y 13, 17 y 19, 29 y 31, 41 y 43, 59 y 61, y 71 y 73. Se conocen muchos otros ejemplos, algunos de valores colosales. Hasta no hace mucho, el ejemplo de máximo valor era una pareja de números primos de 303 cifras, descubierta por Michael A. Penk en 1978. Tal valor

2	3	5	7	11	13	17	19	23	29	31	37	41	43	47	53	59	61	67	71	73	79	83	89
1	2	2	4	2	4	2	4	6	2	6	4	2	4	6	6	2	6	4	2	6	4	6	
1	0	2	2	2	2	2	2	4	4	2	2	2	2	0	4	4	2	2	4	2	2		
1	2	0	0	0	0	0	2	0	2	0	0	0	2	4	0	2	0	2	2	0			
1	2	0	0	0	0	2	2	2	2	0	0	2	2	4	2	2	2	0	2				
1	2	0	0	0	2	0	0	0	2	0	2	0	2	2	0	0	2	2					
1	2	0	0	2	2	0	0	2	2	2	2	2	0	2	0	2	0						
1	2	0	2	0	2	0	2	0	0	0	0	2	2	2	2	2							
1	2	2	2	2	2	2	2	0	0	0	2	0	0	0	0								
1	0	0	0	0	0	0	2	0	0	2	2	0	0	0									

Conjetura de Gilbreath sobre las diferencias absolutas de la sucesión de primos

n	SUCESION	K
1	1,1	2
2	1, 2, 1	3
3	1, 3, 2, 3, 1	5
4	1, 4, 3, 2, 3, 4, 1	7
5	1, 5, 4, 3, 5, 2, 5, 3, 4, 5, 1	11
6	1, 6, 5, 4, 3, 5, 2, 5, 3, 4, 5, 6, 1	13
7	1, 7, 6, 5, 4, 7, 3, 5, 7, 2, 7, 5, 3, 7, 4, 5, 6, 7, 1	19

Triángulo de números primos (Leo Moser)

fue rebasado un año después (en 1979), al ser descubiertos por A.O.L. Atkin y Neil W. Rickert dos pares mucho mayores: $694503810 \cdot 2^{2304} \pm 1$ y $1159142985 \cdot 2^{2304} \pm 1$. En el mayor de estos pares gemelos cada número comienza por 4337..., termina con 17760 ± 1, y está formado por 703 dígitos.

La conjetura sobre pares de primos gemelos puede generalizarse a pares de primos que se diferencien en un número par n dado de antemano. (Dejando aparte el 2, ningún par de primos puede tener diferencia impar, pues de ser así, uno de ellos sería par, y por tanto, compuesto.) Podemos generalizar todavía más la conjetura, estudiando la existencia de ciertas pautas finitas de números primos separados por diferencias pares previamente dadas. Por ejemplo, las siguientes ternas de números primos se ajustan todas al esquema k, $k + 2$, $k + 6$: 5, 7 y 11; 11, 13 y 17; 17, 19 y 23; 41, 43 y 47; 101, 103 y 107.

Se estima que con todos los esquemas de este tipo no prohibidos por consideraciones sobre divisibilidad habrá infinidad de ejemplos. (La pauta k, $k + 2$, $k + 4$ tiene solamente una solución con números primos, a saber, 3, 5, 7; pues toda terna mayor ajustada a esta pauta contendrá un número divisible por 3.) Se cree también que existirán infinitas cuaternas de la forma k, $k + 2$, $k + 6$ y $k + 8$ (el ejemplo mínimo es 5, 7, 11 y 13). Hay pautas para las que se carece de ejemplos, o se dispone de tan sólo uno. R. E. Crandall ha llamado la atención sobre la pauta exhibida por el octeto 11, 13, 17, 19, 23, 29, 31 y 37. Seguramente existirán otros ejemplos de este mismo modelo, pero hasta ahora no se ha encontrado ninguno.

Los números de Mersenne –que son números de la forma $2^n - 1$, o sea, potencias de 2 disminuidas en una unidad– han fascinado a los amantes de la teoría de números desde tiempos clásicos, especialmente por su relación con los números perfectos: números que son suma de sus divisores, incluido el 1, pero no el propio número (6, 28, 496,...). Si un número de Mersenne es primo, automáticamente genera un número perfecto mediante la fórmula de Euclides $2^{n-1}(2^n - 1)$, donde el número contenido en el paréntesis es uno de los primos de Mersenne.

Es fácil demostrar que los números de Mersenne no pueden ser primos a menos que sus exponentes sean también primos. Sabiendo que el exponente n es primo, ¿será primo el número de Mersenne? La ley fuerte de los pequeños números incita a pensar que sí, pues así sucede cuando n es igual a 2, 3, 5 y 7. Mas la ley falla para $n = 11$, porque $2^{11} - 1 = 2047$, que es descomponible en 23×89. La ley vuelve a ser cierta para $n = 13$, $n = 17$ y $n = 19$, pero otra vez es falsa para $n = 23$. De aquí en adelante los éxitos se hacen más y más raros. Por el momento se conocen solamente 27 primos de Mersenne (y por tanto, tan sólo 27 números perfectos). El 27-ésimo primo de Mersenne, $2^{44497} - 1$, fue descubierto en 1979 mediante un programa de ordenador preparado por David Slowinski con la ayuda de Harry L. Nelson, en el Lawrence Livermore Laboratory de la Universidad de California. Tal número empieza por 854..., termina por ...671, y tiene 13.395 cifras. Nadie sabe si el número de primos de Mersenne es finito o infinito y, ni siquiera, si existirá un vigésimo-octavo en la serie.

Los números de Fermat son de la forma $2^{2^n} + 1$. Para $n = 0$, $n = 1$, $n = 2$, $n = 3$ y $n = 4$ tales números son primos (a saber, 3, 5, 17, 257 y 65337). Pierre de Fermat creyó que todos los números de esta forma tendrían que ser primos, pero pasó por alto que al tomar $n = 5$ resulta 4294967297, que se factoriza en producto de 641×6700417, ambos primos. No se han descubierto otros primos de Fermat, aparte de los cinco ya conocidos por él; nadie sabe si existirán o no otros de la misma serie.

Veamos ahora una curiosa pauta que relaciona números primos con números factoriales. El factorial de n, denotado $n!$, es $1 \times 2 \times 3 \times \ldots \times n$. Prestemos atención a la alternativa de signos positivos y negativos de la serie siguiente:

$3!-2!+1!=5$
$4!-3!+2!-1!=19$
$5!-4!+3!-2!+1!=101$
$6!-5!+4!-3!+2!-1!=619$
$7!-6!+5!-4!+3!-2!+1!=4421$
$8!-7!+6!-5!+4!-3!+2!-1!=35899$

En cada uno de estos casos, el número del segundo miembro es primo. Desdichadamente, la ley fuerte de los pequeños números falla en el paso siguiente. Se obtiene 326981, producto de los números primos 79 y 4139. Los casos en que vuelven a resultar números primos corresponden a $n = 9$, $n = 15$ y $n = 19$.

La tabla que vemos en la ilustración inferior de la izquierda se construye como sigue. Comenzamos por 41, y sumamos después 2, obteniendo el número primo 43. A 43 le sumamos 4, y obtenemos el número primo 47. A 47 le sumamos 6, con lo que se obtiene 53, también primo. Prosigamos de esta forma, tomando cada resultado como primer número de la fila siguiente, sumando sucesivamente los números 2, 4, 6, 8,... en todos los casos que vemos en la tabla resulta un número primo. ¿Continuará la progresión indefinidamente, o llegará a fallar en algún momento? Daré la solución el mes que viene.

La curiosidad que se muestra en la ilustración superior de esta misma página fue construida por el matemático canadiense Leo Moser. Examinando la pauta vemos que cada sucesión se deduce de la precedente por inserción de n. o sea, del número de orden de la

NUMEROS PARES	NUMEROS PRIMOS
↓	↓
41 + 2	= 43
43 + 4	= 47
47 + 6	= 53
53 + 8	= 61
61 + 10	= 71
71 + 12	= 83
83 + 14	= 97
97 + 16	= 113
113 + 18	= 131
131 + 20	= 151
151 + 22	= 173
173 + 24	= 197
197 + 26	= 223
223 + 28	= 251
251 + 30	= 281
281 + 32	= 313
313 + 34	= 347
347 + 36	= 383
⋮	

¿Hasta cuándo irán saliendo números primos?

fila, entre todos los pares de números que sumen n. El número k de la derecha denota el número de términos de cada sucesión. Observemos que los seis primeros valores de k son los seis primeros números primos. El siguiente valor de k es 19, saltando el 17, si bien 19 continúa siendo primo. ¿Serán primos todos los valores de k? ¿Hay alguna fórmula que permita calcular el n-ésimo valor de k? Responderé a estas dos cuestiones en la sección del mes próximo.

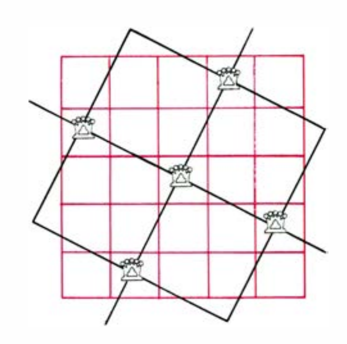
Exceptuando el 2, todos los números primos son de la forma $4k \pm 1$, lo que significa que todos los números primos distintos de 2 difieren en una unidad de un múltiplo de 4. (Que así sucede se comprende fácilmente observando que todo número impar es una unidad más o una unidad menos que un múltiplo de 4.) Escribamos los números primos impares en orden consecutivo, disponiendo los primos de la forma $4k - 1$ en la primera fila, y debajo de los anteriores los de tipo $4k + 1$:

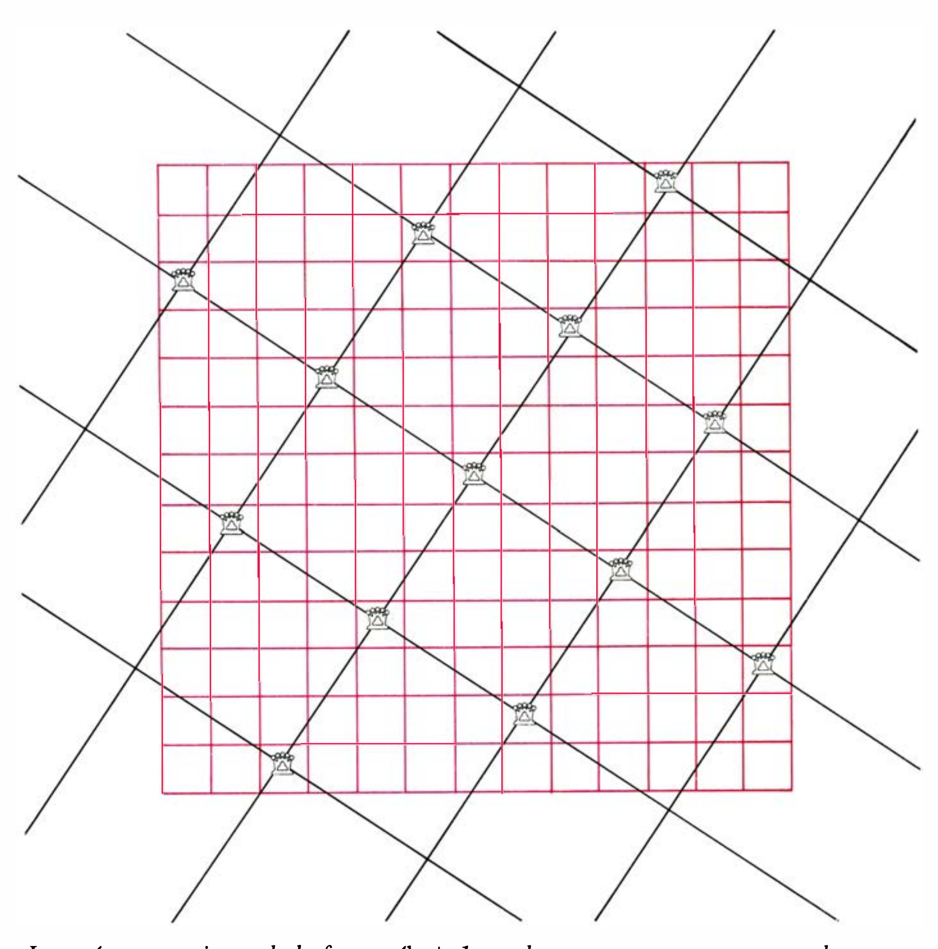
3 7 11 19 23 31 43 47 59 67 71 79 83
5 13 17 29 37 41 53 61 73

Hasta este momento, la hilera superior "va ganando la carrera". Mas, si prosiguiéramos indefinidamente, ¿iría siempre por delante la primera fila? No se debería malgastar el tiempo tratando de dirimir empíricamente la cuestión, nos previene Guy, porque tendríamos que avanzar mucho antes de que la segunda fila tomase la cabeza, e incluso entonces no se habría demostrado nada. El eminente matemático John E. Littlewood, de Cambridge, demostró que las filas van relevándose infinitamente en el puesto de cabeza.

A partir de 5, todos los números son de la forma $6k \pm 1$. Haciendo competir estos dos "caballos", veríamos que también van alternándose indefinidamente en la dirección de la carrera. Se han investigado también otras "mangas" entre números primos, como por ejemplo, la formada por los cuatro caballos $8k \pm 1$, $8k \pm 3$. Aunque dista mucho de haber sido demostrado, casi todos los especialistas opinan que en todas las carreras de este tipo, independientemente del número de caballos, cada potro toma la delantera infinitas veces.

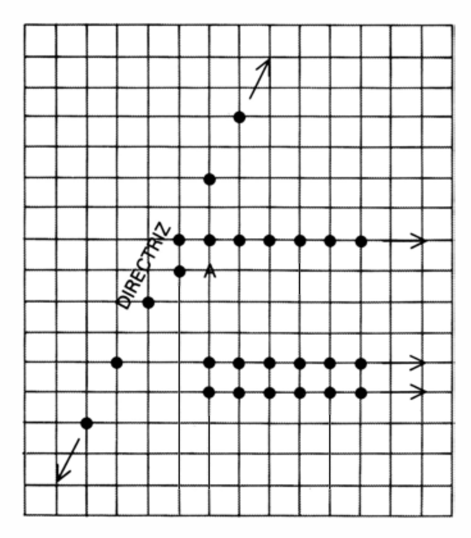
Los números primos de la forma $4k + 1$ (la segunda calle de la carrera $4k \pm 1$) pueden siempre ser expresados como suma de un único par de cuadrados distintos. Así, 5 es igual a $4 + 1$, 13 es igual a $4 + 9$, y de este modo todos los demás. Este resultado fue demostrado por Fermat, llamándose por ello teorema de Fermat, o teorema de los cuadrados. Tenemos aquí un excelente ejemplo donde la ley fuerte de los pequeños números no resulta engañosa, sino que conduce a un teorema verdadero. Desde antiguo se conocían muchos métodos para demostrar este teorema, pero en 1977, Loren C. Larson, del St. Olaf College, de Minnesota, publicó una nueva y deliciosa demostración basada en el conocido problema de situar n reinas en un tablero de ajedrez de n por n, sin que ninguna de ellas se encuentre amenazada por ninguna otra.





Los números primos de la forma $4k \pm 1$ *pueden expresarse como suma de un par de cuadrados distintos*

La figura de la parte superior de la ilustración de esta misma página presenta la mínima solución del problema de las damas que goza de las siguientes propiedades: (1) hay una reina ocupando la casilla central; (2) todas las demás damas pueden ser alcanzadas desde la casilla central mediante un salto de caballo generalizado, consistente en avanzar m casillas en una dirección, y, seguidamente, otras n casillas en dirección perpendicular (siendo m y n enteros diferentes); y (3), la disposición final tiene cuádruple simetría de rotación (es decir, permanece invariable frente a giros centrales de 90 grados de amplitud). La solución de tamaño inmediatamente mayor que reúne todas estas características puede verse en la parte baja de la ilustración: 13 reinas en un tablero de 13 por 13 casillas.



Solución de la parábola "geotaximétrica"

Si prescindimos de la dama que ocupa el centro, en cada solución de este tipo los cuatro cuadrantes del tablero deben contener el mismo número de piezas. Por consiguiente, el número total de damas situadas en él será de la forma $4k + 1$. Larson demuestra entonces que la condición necesaria y suficiente para construir soluciones como las explicadas es que el número de reinas sea primo y de la forma indicada.

En todas las soluciones del problema, el tablero puede dividirse en cuadrados idénticos más pequeños, de la forma que muestran las líneas inclinadas de los dos ejemplos de la página anterior. Si imaginamos ahora que el tablero ha sido transformado en un toro uniendo los bordes superior e inferior y los márgenes derecho e izquierdo, vemos que cada tablero de lado p está formado por p cuadrados inclinados. Dado que el área del tablero es p^2, el área de cada cuadrado pequeño será $\sqrt{p}$. Como p es la hipotenusa de un triángulo rectángulo de lados m y n (las dos componentes del salto de caballo generalizado), resulta del teorema de Pitágoras que p (área del cuadrado construido sobre la hipotenusa) tiene que ser suma de los cuadrados de m y n. Y como p es cualquier número primo de la forma $4k + 1$, resulta que todo primo de este tipo es suma de dos cuadrados distintos. He presentado aquí la demostración de Larson (que se apoya en anteriores trabajos de Georg Pólya) en forma muy esquemática. Pueden consultarse los detalles en su artículo "A Theorem about Primes Proved on a Chessboard" (*Mathematics Magazine,* vol. 50, n.º 2, págs. 69-74; marzo de 1977).

La cuarta y última estrofa del poema de Spalding concluye muy adecuadamente:

¡Salve, números primos,
extraños e improbables!
¡Que por largo tiempo
los cazadores de fórmulas
hayan de cocerse en abastracción,
pacientes consumirse hasta esqueletos!
Permaneced rebeldes, fenómenos molestos,
irreductibles a sistema, a sucesión,
refractarios a pauta
o explicación.

En la narración de mi visita al doctor Matrix el pasado noviembre se me deslizaron cierto número de errores. Al mencionar la famosa calle Istiklal cometí una falta de ortografía, y al referirme al "sector asiático" debí en realidad decir "la Ciudad Vieja". Las palabras con que me despidió el doctor Matrix fueron sencillamente "Gule gule", literalmente, "ve con risa". He de agradecer a varios lectores sus cartas a este respecto.

Había también un par de errores matemáticos. Yo reseñaba allí que el doctor Matrix había dicho desconocer si el conjunto de sin-rep-omirps (números primos sin cifras repetidas y que escritos en orden inverso son también números primos) era finito o infinito. Al revisar mis notas de la entrevista veo que el doctor Matrix tan sólo se refería al conjunto de omirps. Por mi parte, fue una ingenuidad no darme cuenta de que todo omirp de más de 10 dígitos tendrá forzosamente alguna cifra repetida. Además, como me han hecho observar Harvey P. Dale y otros muchos, ningún número con 10 dígitos distintos puede ser sin-rep-omirp, porque los diez guarismos suman 45, y por tanto, cualquiera de sus permutaciones será múltiplo de 9. Según Harvey Dale, el mayor de los sin-rep-omirps es 987653201. No debí haber mencionado el 11939 entre los omirps cíclicos, pues la permutación 19391 es capicúa. Parece como si el único omirp cíclico fuese el de seis cifras que dio el doctor Matrix, aunque seguramente nunca será posible demostrarlo. Y como me hizo observar John Baum, mi recuerdo de una alusión del doctor Matrix a un reciente artículo de Paul Erdös era inexacto. El descubrimiento de Erdös se refiere al número 70, pero el número cuya propiedad describí yo era el 60. Lo que en realidad demostró Erdös es que, si comenzando a partir de n se va escribiendo una sucesión de enteros mayores que n cada uno de los cuales es relativamente primo a todos los enteros anotados con anterioridad, entonces el máximo número para el cual dicha sucesión está formada tan sólo por números primos o cuadrados de números primos es el número 70.

En noviembre, el doctor Matrix pedía otras descomposiciones de un cubo en tres piezas congruentes. Al dar la solución en diciembre dije no conocer nuevas formas (no triviales) de lograr la trisección. Como muchos lectores se apresuraron a hacerme ver, no podía estar más equivocado. John E. Morse envió la solución más general. Sosteniendo un cubo de forma que uno de sus vértices apunte hacia el ojo, y que su perfil sea un hexágono regular podremos apreciar la simetría tridimensional del cubo. Aunque es demasiado complicada para poder exponerla aquí con detalle, tal simetría permite seccionar el cubo de infinitas maneras en tres piezas congruentes entre sí. Las superficies de separación de las piezas pueden ser planas o alabeadas, de formas cualesquiera, y es fácil imaginar complejas trisecciones, donde las piezas se hallen tan trabadas e interpuestas entre sí que resulte del todo imposible separarlas.

Bennett Battaile me escribió proponiendo una alternativa al problema de rotular los vértices de un cubo con los números de 0 a 7, de forma que la suma del par situado en los extremos de cada arista sea primo. Ahora se trata de lograr que la suma sea número compuesto. Desechando las variantes deducidas por giros y simetrías, también este problema resulta tener solución única.

La ilustración que encabeza esta página muestra la solución del problema del mes pasado, consistente en construir una parábola "taximétrica" de foco y directriz dados.

Al comentar los grupos esporádicos el pasado mes de agosto, se me deslizaron un par de errores. Solomon W. Golomb me hizo notar que la famosa carta de Evariste Galois no fue escrita por éste la noche anterior a su muerte, sino la anterior al trágico duelo del 30 de mayo de 1832. Galois permaneció con vida un día más, muriendo de peritonitis el día 31. Joseph Gallian me escribió diciéndome que George Abram Miller no se proponía demostrar la inexistencia del grupo esporádico M_{24}, sino demostrar que el grupo no era quíntuplemente transitivo sobre los enteros de 1 a 24. Su demostración, que resultó ser defectuosa, hubo de ser retirada.



Escarcha helada sobre el vidrio de un ventanal, fotografiada por Gera Dillon

Taller y laboratorio

Estudio sobre las figuras de escarcha helada en las ventanas. Y otro tema: un reloj solar de fácil lectura

Jearl Walker

Para mí, una de las delicias del invierno es la escarcha que adorna los vidrios de mis ventanas, cuyos cristales adoptan múltiples formas y dibujos. Los hay alargados como varillas, otros componen motivos florales. Estos cristales han crecido del agua sobreenfriada que contiene el aire contiguo a la ventana, proceso que inició algún agente nucleador invisible, como una partícula de polvo, partiendo de un monocristal para luego expandirse hasta recubrir una hoja entera del ventanal.

La forma que aparece, de entre las varias posibles, depende principalmente de la temperatura en las proximidades del cristal, pero la relación exacta no se conoce muy bien. Para determinar la forma del cristal puede ser crítica hasta una diferencia de temperatura de pocos grados. Y puesto que pueden observarse formas distintas desarrollándose en una misma ventana, habitualmente a diferentes alturas, debe ocurrir que la temperatura variará al menos unos pocos grados entre los distintos puntos de la ventana. En la página opuesta se adjunta una fotografía de escarcha helada que me ha enviado Gera Dillon, fotógrafo profesional de Morin Heights, Quebec, otro entusiasta de la escarcha en los ventanales.

La estructura cristalina fundamental del hielo es hexagonal. El plano del hexágono se llama plano basal y el eje perpendicular a éste es el eje *c*. Otros tres ejes, denominados *a,* atraviesan el hexágono por sus lados. Observado en cualquier dirección no paralela al eje *c,* el hielo de una ventana es birrefringente. Con este término se expresa el hecho de que la luz puede verse afectada por dos índices de refracción diferentes al atravesar el hielo.

En mi artículo de febrero de 1978 les expliqué las propiedades generales de los materiales birrefringentes. Resumiéndolas brevemente, diremos que los materiales birrefringentes tienen un eje "rápido" y otro "lento"; el índice de refracción depende de cómo esté polarizada la luz al atravesar el material, de forma que el índice es mayor si la polarización es paralela al eje lento, e inferior si la polarización es paralela al eje rápido.

Supongamos que se dirige luz polarizada linealmente (o sea polarizada a lo largo de un solo eje perpendicular al rayo) a través de un material birrefringente y que su eje de polarización forma ángulos no nulos con los dos ejes, lento y rápido, del material. Matemáticamente, la polarización puede separarse en dos componentes, una paralela al eje lento y la otra componente paralela al rápido. Al penetrar en el material, ambas ondas estaban en fase, pero dada la existencia de dos índices de refracción es probable que salgan con una relación de fases diferente.

Como resultado, probablemente la polarización de la luz emergente será distinta de la polarización de la luz incidente. Aunque esta nueva polarización siga siendo tal vez lineal, su eje estará orientado en una nueva dirección. También podría ocurrir que la luz emergente estuviera polarizada circular o elípticamente, lo que significaría que el eje de polarización giraría alrededor del rayo a la vez que éste se propagase.

Si la luz emergente se encuentra con un filtro polarizador, lo atravesará o no según sea la orientación relativa de los ejes de polarización de la luz y del filtro. Los materiales birrefringentes suelen estudiarse colocándolos entre dos filtros polarizadores. Entonces, la luz al atravesar el primer filtro se polariza linealmente y luego, cuando atraviesa el material birrefringente, es muy probable que cambie de polarización, de manera que al encontrar el segundo filtro será transmitida o absorbida según sea su nueva polarización con respecto a la de este segundo filtro.

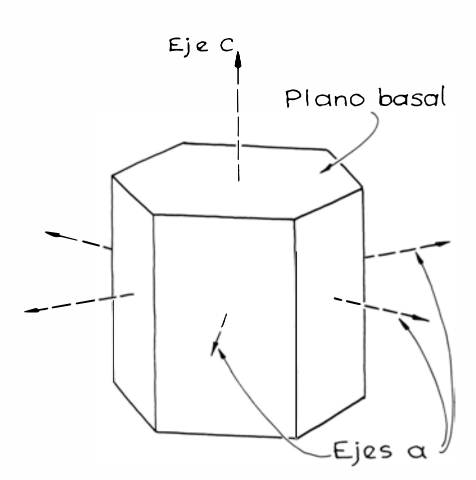
El cambio de polarización de la luz al atravesar el material birrefringente depende de los índices de refracción para los ejes lento y rápido, del espesor del material y de la longitud de onda de la luz. Si se envía luz blanca a través de los dos filtros y del material birrefringente, habrá algunas longitudes de onda que salgan del material polarizadas de forma que puedan atravesar el segundo filtro y ser observadas. Sin embargo, otras longitudes de onda no podrán atravesarlo; y así, aunque se haya lanzado luz blanca sobre el primer filtro, del segundo emergerán sólo determinados colores. Girando cualquiera de los dos filtros o el material birrefringente, mudarán los colores emergentes del segundo filtro.

Ese marco de estudio nos sirve para abordar la escarcha helada sobre una ventana, por la sencilla razón de que el hielo es algo birrefringente. Para ver los colores, basta colocar una lámina polarizadora a cada lado de la ventana tapizada de escarcha. Entonces, todos los cristales de hielo que tengan el espesor y la orientación convenientes darán colores; fenómeno que no sucederá con aquellos otros cuyos ejes *c* sean paralelos a la visual, porque esta orientación elimina la birrefringencia.

Los colores que produce la escarcha pueden verse sin ayuda de filtros polarizadores cuando se haya fundido parte de la misma formando un pequeño charco en la base del marco de la ventana. Ocurre entonces que la luz difusa procedente del firmamento puede estar fuertemente polarizada merced a los mecanismos que les describí aquí en marzo de 1978 y así, si la ventana es iluminada por esta luz, ya no se precisa el primer filtro. Por otra parte, si esta luz atraviesa la escarcha y luego se refleja en el agua, la reflexión puede hacer las veces del segundo filtro, ya que la reflexión puede polarizar la luz. Así pues, al mirar al charquito, se verán versiones coloreadas de las figuras incoloras de escarcha helada existentes en la ventana.

C. K. Sloan, de Arapahoe, Carolina del Norte, ha diseñado un reloj solar de un tipo muy poco corriente. A diferencia de muchos modelos, este "reloj solar analémico" está preparado para corregir los efectos de la declinación del sol y de la longitud y latitud del propio reloj. En él pueden leerse, además, la hora universal y la hora local aparente que indican los relojes solares.

El aparato de Sloan es muy sencillo: un poste vertical proyecta su sombra sobre un cuadrante ideado por él. Aunque en algunas versiones el cuadrante tiene una anchura de varios centímetros, en las de tamaño mayor puede medir hasta casi 25 por 50 metros. Pero en cualquiera de sus versiones, el cuadrante es de gran



Estructura de un cristal de hielo

belleza y trae a la memoria las antiguas representaciones del recorrido del sol a través del firmamento.

Muchos relojes emplean un estilo, que es el objeto que proyecta su sombra, apuntado oblicuamente al polo celeste, o cenit del firmamento situado sobre el polo geográfico. Esta sombra se proyecta sobre una red de líneas rectas, llamada cuadrante, que parten del pie del estilo, de tal forma que cuando el sol se encuentra en su punto más alto, la sombra cae directamente sobre la señal de mediodía, que es una de las rectas que parten radialmente del pie del estilo.

Si bien éstos relojes solares en su versión para jardín son muy atractivos, la determinación de la hora con ellos obliga a recurrir a tablas de conversión y hay que dedicarles algo de trabajo. El cuadrante está armonizado con el paso del sol, no engranado a ningún mecanismo de relojería. Así, cuando el sol se encuentra en su punto más alto, la hora local aparente es mediodía y cuando el sol atraviesa un arco de 7,5 grados en el firmamento es que ha transcurrido media hora.

El procedimiento, que es directo, presenta como mínimo dos inconvenientes. Uno de ellos es que la hora local aparente depende directamente de la longitud geográfica del lugar en que está emplazado el reloj solar. (Dos relojes solares situados en longitudes diferentes indicarán horas distintas.) El otro es que la duración de una hora de tiempo local aparente cambia en el transcurso del año a causa de la forma de la órbita terrestre alrededor del sol. Como resultado, la velocidad aparente del sol a través del cielo varía en el transcurso de los meses. Cada hora de tiempo local aparente permanece proporcional al movimiento del sol a través de un arco de 15 grados, y de esta forma varía la duración de una hora.

Hay otro sistema horario que emplea la llamada hora local media y que es similar al de la hora local aparente, salvo en que está armonizado con el paso de un sol ficticio que se mueve uniformemente a través del cielo durante todo el año. Este movimiento uniforme implica que la duración de una hora no se altere con el transcurso de los meses. La diferencia entre ambos sistemas horarios se llama ecuación de tiempo (antiguamente la voz *ecuación* tenía a veces el significado de *error*). Esta diferencia puede encontrarse tabulada en los libros de consulta dedicados al cronometraje de precisión de estrellas y planetas.

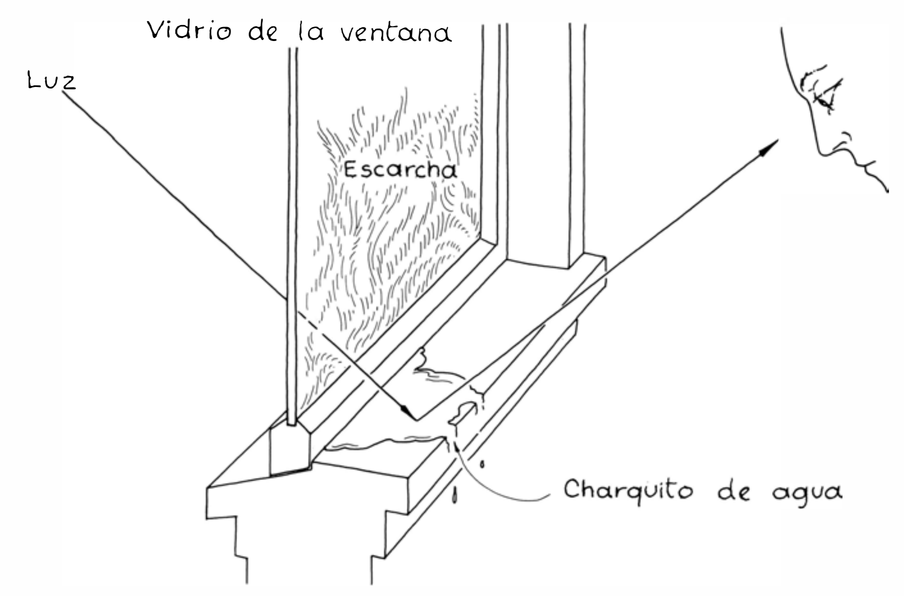
Durante dos períodos del año se dice que un reloj solar se atrasa porque la hora que señala está rezagada respecto a la hora local media. A lo largo de los dos períodos restantes, el reloj solar se adelanta. Durante el período de retraso, para convertir la hora de un reloj solar en hora local media debe sumarse la ecuación de tiempo, correspondiente al día concreto, a la hora local aparente señalada por el reloj. Durante los períodos de adelanto, de la hora local aparente deberá restarse la ecuación de tiempo. Este proceso requiere, por supuesto, un acceso inmediato a una tabla de la ecuación de tiempo. Y lo que es más, el observador seguirá frente al inconveniente de que la hora local media es sólo local.

Habitualmente, la hora que marca un reloj es la hora "oficial" pensada para "aprovechar la luz". Como es bien conocido, el mundo está dividido en husos horarios en los que, cualquiera que sea la longitud de cada reloj, cada uno de ellos mantiene la hora local media correspondiente a un determinado meridiano que pasa por el huso. Este meridiano particular es el meridiano oficial del huso horario. El sistema tiene la ventaja de que todos los relojes incluidos en el mismo huso marcan la misma hora.

Para pasar la hora leída directamente en un reloj solar (hora local aparente) a la hora leída en un cronómetro (hora oficial) se requieren dos operaciones. Primero, la hora del reloj solar debe pasarse a hora local media sirviéndonos de una tabla de valores de la ecuación de tiempo. Después, la hora local media relativa al lugar del reloj solar debe convertirse a la hora oficial correspondiente a su huso horario.

Esta segunda conversión equivale a determinar la diferencia de longitudes entre el reloj solar y el meridiano oficial del huso horario. Si el reloj está situado al oeste del meridiano oficial, su hora local media estará retrasada. Así, un cronómetro señalará mediodía cuando el sol ficticio pase por el meridiano oficial, pero el sol ficticio pasará más tarde por encima del reloj solar. Por consiguiente, la hora local media indicada por el reloj solar se retrasa respecto a la del cronómetro. Si el reloj solar se encuentra al este del meridiano oficial, su hora local media estará adelantada respecto a la del cronómetro. Cuanto mayor sea la diferencia entre las longitudes del meridiano oficial y del reloj solar, tanto mayor será su discrepancia (cuatro minutos de hora por cada grado de diferencia en longitud).

Así pues, tratar de leer la hora real en un reloj solar común puede resultar una tarea bastante tediosa. El diseño de Sloan permite al observador leer la hora real directamente en el reloj solar sin recurrir a tablas ni a cálculos. Todas las correcciones están incorporadas en el diseño mismo. En vez de estilo



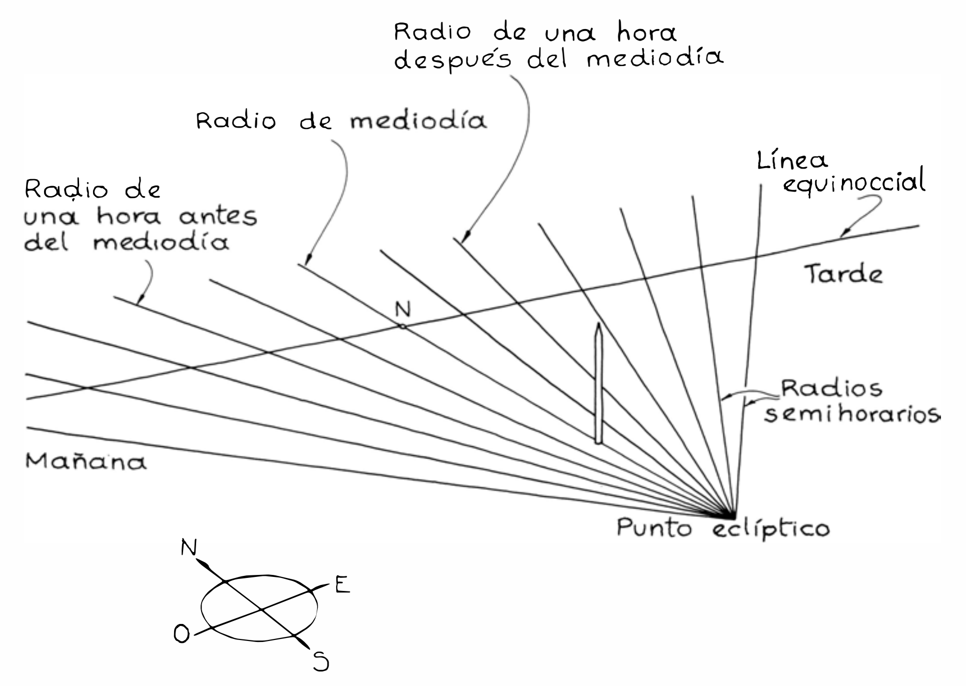
Cómo ver colores en una figura de escarcha helada

oblicuo y seguir la rotación de su sombra en torno al pie, Sloan emplea un estilo vertical y sigue la punta de su sombra. En el transcurso del día, la punta de la sombra se desplaza a través del cuadrante dispuesto en torno al poste y en el cual están incorporadas las correcciones correspondientes a la ecuación de tiempo y a la distancia del reloj solar al meridiano oficial.

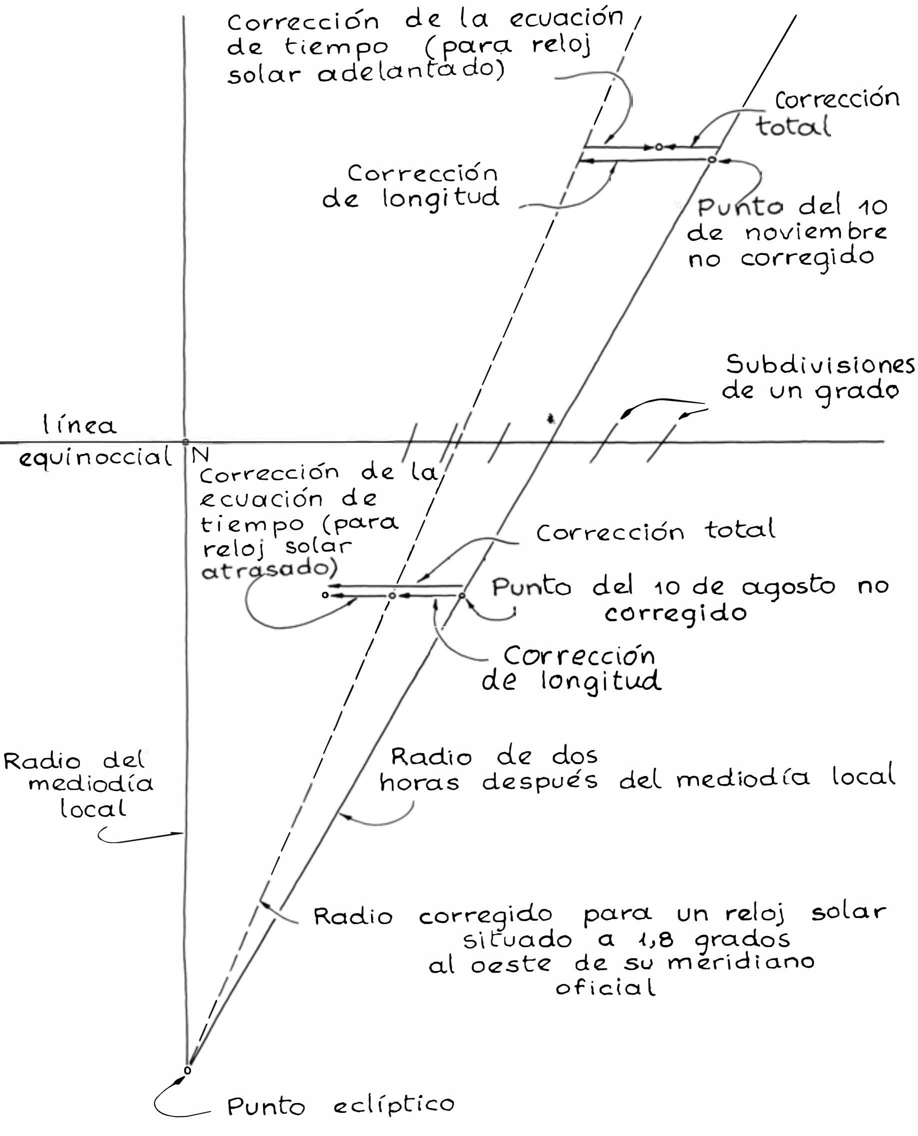
Normalmente, la trayectoria de la punta de la sombra de un estilo vertical sobre una superficie horizontal no se traslada directamente de oeste a este a medida que el sol va cruzando el firmamento. Para determinar esta trayectoria, Sloan erigió un estilo vertical y marcó dos puntos claves en su proximidad. En dirección al sur del estilo, a una distancia igual al producto de la altura del estilo por la cotangente de la latitud del reloj solar, marcó lo que el llamó el punto eclíptico. A partir de éste había que trazar las rectas radiales, una por cada media hora.

Con tal fin determinó otro punto (señalado por *N* en la ilustración superior de esta misma página) en dirección norte del estilo y a una distancia igual al producto de la altura de éste por la tangente de la latitud. Por este punto trazó una línea recta, que llamó línea equinoccial, extendiéndola hacia el este y el oeste. Esta recta encierra la particularidad de que en los dos equinoccios (el 21 de marzo y el 23 de septiembre) la punta de la sombra del estilo se desplaza a lo largo de ella y al mediodía de esas fechas, cuando el sol está directamente encima, la punta de la sombra cae en *N*.

En 30 minutos, la tierra gira 7,5 grados y la punta de la sombra se desplaza por la línea equinoccial una longitud igual al producto de la tangente de 7,5 grados por la secante de la latitud del reloj de sol y por la altura del estilo. Sobre la línea equinoccial Sloan efectuó divisiones equivalentes a intervalos de media hora hasta cinco horas, antes y después del punto de mediodía. Seguidamente trazó rectas radiales a partir del punto eclíptico (el situado al sur del estilo) hasta dichas señales semihorarias. Llamaré, a las rectas, radios semihorarios. Con esto, a Sloan ya le era posible leer la hora en su reloj solar durante dos días al año: los correspondientes a los equinoccios. En esas jornadas, la punta de la sombra empezaba tocando el extremo oeste de la línea equinoccial y luego se desplazaba a lo largo de esta recta, pasando sobre las señales semihorarias hasta alcanzar el punto de mediodía. A partir de aquí, continuaba por la recta en dirección este.



Cuadrante básico de un reloj solar



Corrección de las posiciones de los puntos de sombra

Pensé en un principio que la sombra se desplazaría de la misma forma durante el resto de los días del año; pero no es así. En los demás días, la punta de la sombra cae al norte o al sur de la línea equinoccial, según la estación. Para ser más exactos, su distancia a la línea equinoccial depende de la altura máxima del sol durante el día. Esta altura, corrientemente llamada declinación del sol, se toma como cero en los equinoccios. En otoño e invierno, la declinación es negativa porque la altura máxima del sol en el firmamento es relativamente baja. En primavera y verano la declinación es positiva pues la altura máxima es relativamente alta.

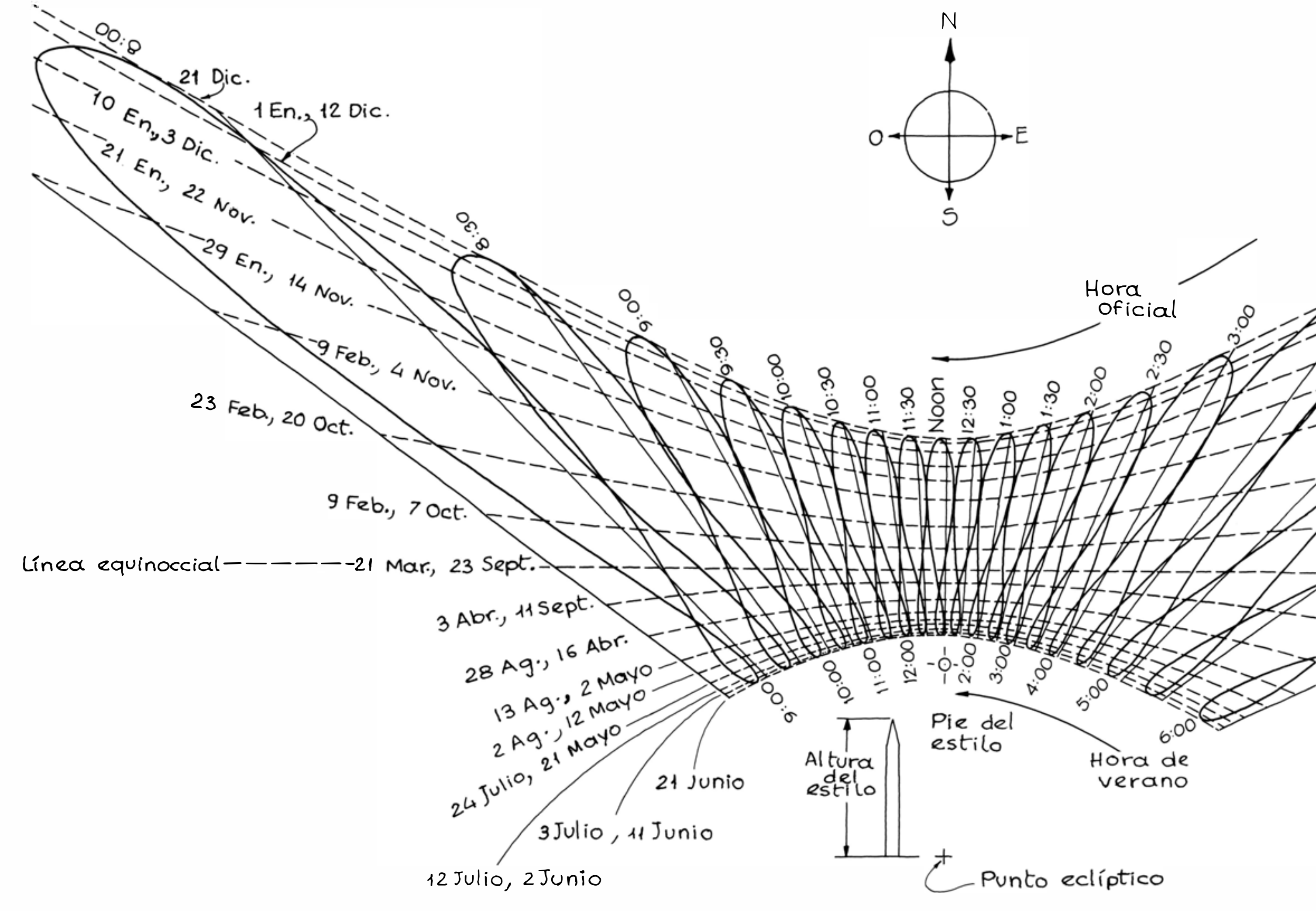
En cada equinoccio, la punta de la sombra cae en el punto *N* de la línea equinoccial cuando el sol alcanza su punto más alto en el firmamento. En otoño e invierno, como la declinación es menor que cero, el sol está más bajo en el firmamento cuando pasa por el punto de mediodía, y por ello la punta de la sombra cae al norte de la línea equinoccial. En primavera y verano la sombra cae al sur de la línea equinoccial y más cerca del estilo.

Los cálculos para determinar las posiciones de mediodía de la punta de la sombra son sencillos. La distancia al estilo se halla multiplicando la altura de éste por la tangente de un ángulo igual a la diferencia algebraica de la latitud menos la declinación del sol. En primavera y verano, se sustituye en la expresión un valor positivo de la declinación, que se resta de la latitud. En otoño e invierno se sustituye un valor negativo, lo que significa que en realidad, y en virtud de la neutralización de los dos signos negativos, a la latitud se suma el valor absoluto de la declinación.

Aunque de esta forma se determinan con gran facilidad los puntos de mediodía de la sombra, no ocurre lo mismo para los puntos semihorarios correspondientes al resto del día. En los equinoccios las señales semihorarias caen todas en la línea equinoccial, que es una recta, pero en los demás días del año la sombra describe una trayectoria curva. En otoño e invierno la sombra comienza el día por el noroeste, se desplaza a lo largo de una curva hacia el sudeste hasta llegar al punto de mediodía correspondiente a ese día, para dirigirse luego hacia el noreste por una curva simétrica con su trayectoria matutina. En primavera y verano, la trayectoria comienza en el sudoeste, se desplaza hacia el noreste hasta llegar al punto de mediodía y luego se curva hacia el sureste.

Sloan ha deducido fórmulas que permiten calcular la posición de la punta de la sombra para cualquier hora del día y cualquier día del año. Con los resultados obtenidos le fue posible trazar en torno del poste un cuadrante que indicase la hora local aparente. Yo he redactado un programa de ordenador para efectuar los cálculos relativos a un reloj solar situado en el hemisferio norte, que puede modificarse fácilmente para aplicarlo a otro situado en el hemisferio sur. El programa, que se representa en la ilustración de dos páginas más adelante, está escrito en el lenguaje de ordenador llamado Basic de nivel II para un ordenador personal Radio Shack TRS-80.

El programa calcula las posiciones de los puntos de sombra a lo largo de los radios semihorarios que parten del punto eclíptico. La línea 10 contiene la



Abaco de un "reloj solar analémico" de C. K. Sloan

latitud del reloj, que yo tomé como 35 grados, y la altura del estilo, que supuse de un metro. (La cantidad "RPD" convierte grados en radianes, para posteriores exigencias de las funciones trigonométricas del programa.) Al llegar a la línea 20, el programa pide que se introduzca una fecha y la declinación del sol (en grados) en dicho día. Entre las líneas 80 y 120 se calculan las posiciones de los puntos de sombra para 10 instantes antes del mediodía y para el mismo mediodía. Como los resultados serán los mismos para los instantes equivalentes posteriores al mediodía, de hecho quedan calculadas las posiciones correspondientes a 21 instantes del día. (Algunos resultados parecerán erróneos si se olvida que en invierno el sol sale más tarde y se pone antes.)

La línea 130 presenta en la pantalla del ordenador la hora (horas respecto al mediodía) y la distancia (en metros) de las posiciones de los puntos de sombra. Por comodidad, al representar los resultados la distancia se da respecto a la línea equinoccial y no respecto al punto eclíptico. Por ejemplo, supongamos que en el instante correspondiente a dos horas antes del mediodía, es decir, cuando la hora local aparente señala las 10 de la mañana, la distancia impresa en la pantalla es de 0,2 metros. Marcaré este valor en el radio correspondiente a las 10 horas de tiempo local aparente y el punto estará a 0,2 metros de distancia de la línea equinoccial y en el lado norte. Un punto igual deberá indicarse en el radio correspondiente a dos horas después del mediodía, puesto que los ábacos de mañana y tarde son simétricos. Si la distancia señalada por el ordenador fuese negativa, el punto habría de marcarse al sur de la línea equinoccial.

Ejecuté el programa para obtener los datos de 34 días repartidos regularmente de manera aproximada a lo largo del año y, entre ellos, incluí los equinoccios (cuando la declinación es nula) y los solsticios de invierno y verano (cuando la declinación es mínima o máxima). Siguiendo las instrucciones de Sloan, elegí los días de forma que se lograran 17 declinaciones diferentes, y así cada fecha tuviera una pareja para la cual la declinación del sol fuese la misma. Por ejemplo, el 14 de noviembre y el 29 de enero tienen asociados los dos la misma declinación de −18 grados aproximadamente.

Utilizar estos valores para trazar las posiciones de los puntos de sombra sería trivial por sí mismo. Así, para un momento cualquiera del día, por ejemplo las dos de la tarde de hora local aparente, se marcarían las distancias en el correspondiente radio que parte del punto eclíptico, de forma que el punto representante del 21 de junio estaría más cercano al punto eclíptico que el del 21 de diciembre, y los demás se marcarían en distintos lugares entre los dos mencionados, pero todos ellos sobre el mismo radio semihorario que parte del punto eclíptico. Entonces, el ábaco resultante sería similar al de un reloj solar corriente.

La eficacia del reloj solar de Sloan se pone de manifiesto cuando las posiciones de los puntos sufren dos correcciones. Primero, se convierten a hora local media corrigiéndolas con la ecuación de tiempo. Después, para pasarlas a hora oficial, se les corrige la diferencia de longitudes entre la del reloj y la del meridiano local oficial. Ambas correcciones equivalen a un giro de los puntos alrededor del punto eclíptico. Este giro lo calcula el programa, con ambas correcciones incluidas.

La segunda corrección es sencilla. Si el reloj se encuentra al oeste del meridiano oficial, su hora local media estará retrasada respecto a la hora local media del meridiano; si se encuentra al este, su hora local media estará adelantada. El ajuste de la longitud se efectúa en la línea 10 del programa, por donde se introduce la diferencia de longitud "DL". Yo he supuesto un reloj situado a 1,8 grados al oeste del meridiano oficial. Para relojes situados en otros lugares, habrá que cambiar esta cifra. Si el reloj está al oeste del meridiano oficial, la diferencia de longitudes se introduce como número positivo, y, si está al este, como número negativo.

La corrección de longitud no reviste especial complicación porque es la misma para todos los puntos del cuadrante. En el caso que he puesto como ejemplo de un reloj a 1,8 grados al oeste de su meridiano oficial, todos los puntos del cuadrante girarían alrededor del punto eclíptico un ángulo de 1,8 grados hacia el oeste. El aspecto del cuadrante permanecería inalterado, puesto que todos los puntos representantes de una hora determinada seguirían estando sobre una misma recta que pasa por el punto eclíptico. La única diferencia tras esta corrección es que esta recta ya no será uno de los radios semihorarios.

Cuando se efectúa la otra corrección, que gracias a la ecuación de tiempo pasa la hora local aparente del reloj a su hora local media, el aspecto del ábaco cambia totalmente. La línea 30 convierte la ecuación de tiempo en un ángulo: 0,25 grados por cada minuto de tiempo. Este ángulo y la diferencia de longitudes se presentan en la parte superior de la pantalla del ordenador y luego se combinan y aparece el ángulo de corrección total. Si éste es positivo, los puntos deben girarse hacia el oeste alrededor del punto eclíptico y, si es negativo, el giro debe ser hacia el este.

Una vez girados los puntos de acuerdo con el ángulo de corrección total, el aspecto del ábaco habrá cambiado porque la corrección es distinta para cada uno de los días del año. (La diferencia diaria procede de la ecuación de tiempo, no de la diferencia de longitudes que permanece igual todo el año.) Por consiguiente, el punto representante de un instante determinado de un día dado se correrá en un sentido a partir del radio semihorario, mientras el representante del mismo instante de otro día se correrá a distancia diferente e incluso en sentido contrario.

El resultado es inesperado y bello a la vez. Si se traza una línea que pase por todos los puntos representantes de una hora cualquiera del día se verá un número ocho deformado, o sea, un analema. Con esta línea puede determinarse la hora sobre el reloj solar de

Sloan. Consideremos el analema correspondiente a la línea de mediodía, que se extiende hacia el norte del punto eclíptico. Cuando la punta de la sombra pasa por la línea de mediodía, la hora local aparente es mediodía. Cuando la punta de la sombra pasa por encima de la porción de analema correspondiente a la fecha de ese día, la hora oficial (la señalada por un cronómetro) es mediodía.

Los radios que parten del punto eclíptico tienen un analema cada uno. (Si se construye un cuadrante pequeño, el ábaco puede resultar confuso. Entonces debe trazarse un analema en torno a los radios correspondientes a una hora completa.) La hora local aparente se lee cuando la sombra pasa por un radio determinado, como el correspondiente a las dos de la tarde de hora local. Cuando la sombra pasa por encima de la porción de analema asociado a la fecha de ese día, la hora oficial señala las dos de la tarde.

Para construir otro reloj solar analémico hay que empezar cambiando los valores de la latitud y de la diferencia de longitudes que se reseñan en la línea 10 de mi programa de ordenador. (Se emplea un valor positivo para el oeste y uno negativo para el este; ambos se introducen en radianes.) Además, se sustituye el valor *H* por el correspondiente a la altura del nuevo estilo (en metros). Se introduce el programa. Cuando el ordenador lo pida, se incorpora una fecha y la declinación del sol (en grados) para ese día. Se introduce además la ecuación de tiempo (en minutos) para esa fecha y se le dice si el reloj solar se adelanta o se retrasa respecto a la hora local media. En el delicioso libro sobre relojes solares de Albert Waugh que se cita en la bibliografía final pueden encontrarse tablas de valores medios de la declinación y de la ecuación de tiempo. Valores más exactos pueden recogerse en publicaciones periódicas especializadas, como en el Anuario del Observatorio Astronómico de Madrid.

El programa presenta en la pantalla la diferencia de longitudes (en grados), la ecuación de tiempo (pasada a grados) y el ángulo total que deben trasladarse los puntos en esa fecha. Ofrece, a renglón seguido, la fecha y dos columnas de cifras: una de ellas con la hora (respecto al mediodía) y la otra con la posición del punto correspondiente al norte (valores positivos) o al sur (negativos) de la línea equinoccial.

Para trazar el cuadrante se marcan el punto eclíptico y la línea equinoccial. Luego se señalan los puntos semihorarios a lo largo de la línea equinoccial y se trazan los radios que pasan por cada uno desde el punto eclíptico. Con esto quedará terminado el cuadrante básico; es el momento de introducir las modificaciones de Sloan.

La distancia entre radios representa media hora, o 7,5 grados de rotación (aparente) del sol alrededor de la tierra. A lo largo de la línea equinoccial y entre radio y radio se señalan separaciones de un grado, cada una de las cuales tendrá una longitud igual al producto de la tangente de un grado por la secante de la latitud del reloj solar y por la altura del poste. Estas señales adicionales sirven para facilitar la corrección de las posiciones de los puntos de sombra.

Se ejecuta el programa para obtener los datos de 34 días del año. Cada pase proporcionará un conjunto de puntos de sombra a llevar sobre el cuadrante, un punto por cada radio semihorario. Supongamos que para el radio correspondiente a las dos de la tarde la distancia es 0,2 metros al norte de la línea equinoccial y que el ángulo total sea dos grados hacia el oeste. El punto correspondiente debe trazarse sobre una recta que pase por el punto eclíptico girado dos grados hacia el oeste a partir del radio de las dos de la tarde. Las separaciones de un grado de la línea equinoccial se emplearán para hallar la ubicación correcta. Sobre la recta girada (no dibujada) el punto de sombra está a 0,2 metros al norte de la línea equinoccial. Este punto se rotula con la fecha. Todos los puntos de sombra que se tracen para dicho día estarán corridos el mismo ángulo respecto a los radios principales, por lo que su representación gráfica será bastante rápida.

Una vez trazados todos los puntos correspondientes a los 34 días, se dibuja un analema para cada uno de los radios semihorarios y con ello quedará finalizado el cuadrante para un reloj solar de Sloan. La hora local aparente se lee en los radios semihorarios, como se haría en un reloj solar corriente. La hora oficial se lee en el analema superpuesto a cada radio. Si está vigente la hora de verano, hay que cambiar en una hora la hora oficial.

Los cálculos pueden efectuarse con una calculadora de bolsillo, pero llevarán mucho más tiempo. Los lectores expertos en los lenguajes de ordenadores personales pueden perfeccionar mi programa. Si se pretendiese diseñar relojes solares analémicos para lugares diversos, deberán almacenarse en el ordenador los valores de la declinación y de la ecuación de tiempo para eliminar la necesidad de responder con los datos cuando el ordenador lo precise.

El aspecto del ábaco varía con la latitud de una forma interesante. En lugares cada vez más próximos al ecuador, los radios semihorarios se van acortando porque el sol está, como se sabe, más alto en el cielo y el analema superpuesto a cada radio se convierte en un número ocho más perfecto, perpendicular a la línea equinoccial. En el polo norte el ábaco sería muy diferente. (Y por supuesto sólo serviría en los meses de primavera y verano, ya que el sol no es visible el resto del año.) Este ábaco sería circular, con su radio de mediodía a uno de los lados y el correspondiente a medianoche en el otro y el analema compuesto por bucles sencillos y no dobles.

Es muy probable que ya desde hace mucho tiempo se construyeran relojes solares similares al de Sloan marcando pacientemente la posición de sombras y sin recurrir a cálculos. Personalmente creo que sería divertido imitar a los antiguos construyendo un reloj solar gigantesco, que rivalice por su tamaño con los megalitos de Gran Bretaña y Francia. Verosímilmente, el cuadrante podría hacerse suficientemente preciso para que incluyera las variaciones anuales de la declinación del sol y de la ecuación de tiempo.

```
 1Ø CLS : RPD = .Ø1745 : L= 35.Ø * RPD : DL = 1.8 : H = 1.Ø
 2Ø INPUT "DECLINATION (DEG) ="; DECL : INPUT "DATE ="; A$
 3Ø INPUT "EQUATION OF TIME (MIN) ="; ET : ET = ET * Ø.25
 4Ø INPUT "IS DIAL FAST OR SLOW"; B$ : D = DECL * RPD
 5Ø IF B$ = "FAST" THEN ET = -ET
 6Ø PRINT "DL ="; DL, "EQ TIME ="; ET, "TOTAL ANGLE ="; DL + ET
 7Ø PRINT " ": PRINT "TIME (HRS)", "DIST (METERS)", A$
 8Ø X = TAN (L) : F = 1/ COS (L) : G = X + 1/X : W = 1/ SIN (L)
 9Ø FOR K = Ø TO 1Ø : C = K * 7.5 * RPD : B = TAN (C)
1ØØ E = SQR ( (B * F) ↑ 2 + G ↑ 2) / G : J = 1/ (E * X)
11Ø A = ATN ( (G * E - J) / SQR (W ↑ 2 - J ↑ 2) )
12Ø Z = TAN (A - D) / TAN (A) : DIST = H * (Z - 1) * (G * E - J)
13Ø PRINT T, DIST : T = T + .5 : NEXT K : END
```

Programa de ordenador para calcular las posiciones de la punta de la sombra en un reloj solar situado en el hemisferio norte

Libros

Filosofía del lenguaje, reumatología y belleza de los objetos científicos

Juan José Acero, F. J. García-Conde y Luis Alonso

PRINCIPIOS DE FILOSOFIA DEL LENGUAJE 1, por J. Hierro Sánchez-Pescador; Alianza Editorial, Madrid 1980. Este primer volumen versa, en concreto, sobre los temas de la teoría de los signos, la teoría de la gramática y la epistemología del lenguaje. Con estos *Principios* trata de mostrar las aportaciones más características e iluminadoras que en cuanto al tema del lenguaje se han hecho desde las tendencias filosóficas occidentales vigentes en nuestro siglo: las de carácter especulativo, analítico y dialéctico. En correspondencia con estas líneas del quehacer filosófico, el autor distingue otras tantas concepciones o teorías –en un sentido laxo de este término– del lenguaje: la teoría trascendental, la teoría formal y la teoría social del lenguaje. De acuerdo con la primera de ellas, del lenguaje se subraya su función en la constitución de la conciencia y el que incorpore una interpretación de la realidad; propio de la segunda es considerar al lenguaje como un medio de expresión al que subyace una estructura matemática emparentada con el género de cálculos y sistemas que los lógicos vienen investigando desde finales del siglo pasado; en la teoría social del lenguaje, por su parte, se incide en temas como el de la relevancia del lenguaje para la transmisión de las ideologías o su concepción como trabajo y como mercado. En la presente obra, y salvo muy escasas páginas de los capítulos primero y tercero, la teoría formal del lenguaje asume el papel protagonista. Por ello, no es éste el momento de juzgar hasta qué punto el profesor Hierro ha cumplido sus objetivos; habrá que esperar para ello a la aparición de las subsiguientes partes de estos *Principios*.

Con todo el interés que tiene este proyecto, la forma de exponerlo de su autor es muy modesta. Este primer volumen dista de poseer un carácter expositivo puro y simple; incluye aportaciones de muy diversa índole a los temas que se tratan, tanto bajo la forma de evaluaciones y, ocasionalmente, de muy severas críticas de argumentos o de tesis hoy por hoy notorios, cuanto bajo la de propuestas y sugerencias sustantivas. Esto se hace especialmente manifiesto en el capítulo segundo, contiene una nueva taxonomía de los signos, en el capítulo tercero, con su defensa de la terna 'sistema'/'norma'/'habla', de Eugenio Coseriu, como claro y operativo marco conceptual con cuya ayuda situar y esclarecer conceptos de la filosofía del lenguaje, como el de lenguaje de clase, poco frecuentes en las exposiciones de estos temas, y en el capítulo quinto, con su pormenorizado análisis de la hipótesis, propuesta por Noam Chomsky, de que la especie humana lleva innata la facultad del lenguaje (hipótesis del innatismo lingüístico). Junto a esto, la obra manifiesta también las virtudes de los otros escritos del profesor Hierro: su claridad expositiva, su atención a las fuentes y temas de importancia, su minuciosidad en los planos de la argumentación y la crítica y su independencia de criterio.

Este primer volumen de los *Principios* del profesor Hierro consta de cinco capítulos. En el primero de ellos ("Filosofía del lenguaje: esbozo de justificación") se expone la concepción de la filosofía que tiene el autor en tanto que "interpretación que aspira a ser total, razonada y autónoma" (p. 13), sitúa a la filosofía del lenguaje dentro de esta forma de interpretación y traza sus relaciones con disciplinas científicas como la psicología y la lingüística. En el capítulo segundo ("Signos, signos, signos") se introduce y precisa la noción tradicional de signo –un signo es "todo cuanto representa a otra cosa en algún respecto para alguien" (p. 32)–, se distinguen los elementos del signo y se propone una clasificación de los signos que, aparte su interés terminológico, es novedosa por la forma en que articula los distintos criterios que acoge. Se nos presenta también el concepto de lenguaje como "pluralidad de signos de la misma naturaleza cuya función primordial es la comunicación entre organismos" (p. 38), y se acaba discutiendo la teoría de los signos de Guillermo de Occam y la vetusta idea de un lenguaje de la mente, en la que el profesor Hierro ve un precedente importante de algunas de las reflexiones filosóficas más características de Bertrand Russell, Ludwig Wittgenstein y Noam Chomsky. El capítulo tercero ("¿De lo concreto a lo abstracto o de lo abstracto a lo concreto?") plantea la cuestión metodológica de la relación entre las teorías gramaticales y los datos que las confirman. El capítulo contiene un análisis detallado de los conceptos chomskianos de competencia y actuación lingüística y de las tesis de que la teoría gramatical debe investigar la competencia lingüística del hablante-oyente ideal (el conocimiento que de su propia lengua tendría tal sujeto) y de que los hablantes de una lengua hacen un uso creativo de ella. En lo que hace a la primera de ambas, la actitud del profesor Hierro es decididamente contraria al proyecto de psicologizar la teoría lingüística: "... el mero estudio del sistema lingüístico en cuanto realidad abstraída a partir del comportamiento no obliga a colocar el sistema en ningún sitio, y en particular a situarlo precisamente en la mente" (p. 59). En cuanto a la segunda, concluye su autor que la forma en que Chomsky ha formulado el aspecto decisivo de la creatividad lingüística "deja a este concepto sumido en una vaguedad total e injustificada" (p. 65). El capítulo cuarto ("*Ars Grammatica*") contiene una muy útil puesta al día en temas tales como los universales lingüísticos, o rasgos comunes a todas las lenguas humanas, la concepción generativa de la gramática (es decir, de la gramática concebida como teoría axiomática cuyos teoremas son las oraciones bien formadas de la lengua en cuestión), con atención preferente a las publicaciones de Chomsky y de Jerry Katz, y de los principios y resultados básicos de la teoría matemática de las gramáticas que son de importancia para la lingüística en tanto que ciencia empírica. El capítulo quinto ("Ideas nonnatas") vuelve en lo básico a los temas que el profesor Hierro trató detenidamente en su libro *La teoría de las ideas innatas en Chomsky*. (Véase INVESTIGACION Y CIENCIA, junio de 1977, pp. 108-110.) Se reafirma en sus conclusiones de entonces, como cuando afirma que la "vaporosa psicolingüística chomskiana tiene el peligro, ya suficientemente comprobado, de sumirnos en el reino de las fantasías" (pág. 152). Pero, al margen de que el profesor Hierro haya escrito ya sobre algunas de estas cuestiones, este postrer capítulo se halla notablemente enriquecido tanto en sus re-

ferencias como en su contenido, y para su comprensión basta con seguir el hilo de las partes correspondientes de este primer volumen.

Tras estas referencias al contenido del primer volumen de estos *Principios,* haré a continuación un par de observaciones críticas sobre el apartado más polémico de esta obra: el relativo al rechazo de la epistemología del lenguaje de corte chomskiana. Según el profesor Hierro, Chomsky posee "ideas muy confusas sobre las características metodológicas y epistemológicas de su propia actividad" (p. 59). Eso se haría especialmente claro en lo concerniente a la hipótesis del innatismo lingüístico. Me parece a mí que la mayor parte de las personas interesadas por estos temas estarán de acuerdo en reconocer que esta hipótesis no es, por el momento, susceptible de confirmación. Pero también me parece que el aventurarla y el defender su pertinencia, e incluso su confirmabilidad en principio, no son síntoma de ninguna confusión metodológica. No es ésta la actitud del profesor Hierro; para él, aceptar la teoría generativa de la gramática, en cualquiera de sus versiones, "no obliga... a mantener una posición innatista ni mentalista sobre la adquisición del lenguaje" (p. 151). Esto es cierto, sin duda. Pero a una actitud así se le puede recriminar legítimamente el arrinconamiento de ciertas cuestiones, aunque sean cuestiones que plantean interrogantes para los que carecemos de una respuesta definida. Desligar las investigaciones gramaticales de las investigaciones psicolingüísticas sobre la adquisición del lenguaje, como defiende el profesor Hierro, equivale a exigirles a nuestras gramáticas que pasen únicamente por el tamiz de la adecuación observacional, es decir, exigirles tan sólo que incluyan, de una forma reducida, todo lo que incluiría una cinta magnetofónica en la que se registrase sólo fragmentos del habla de una comunidad lingüística. Aceptemos eso momentáneamente. Lo cierto es que, una vez arribados a alguna propuesta concreta sobre los rasgos de nuestras gramáticas, es inevitable preguntarse por qué poseen esos rasgos precisamente. Y esta pregunta, que nos pide explicaciones sobre la naturaleza de las gramáticas elaboradas, es la que nos conduce a una hipótesis como la innatista, al cifrar en nuestra dote genética y en el proceso de adquisición del lenguaje la clave de la cuestión. La búsqueda de principios que expliquen leyes o hipótesis es, sin embargo, una maniobra metodológicamente impecable.

El recelo del profesor Hierro para con la epistemología del lenguaje chomskiana deriva más bien –me atrevo a aventurar– de su concepción de fondo de lo que para él es el lenguaje humano –un *instrumento* de comunicación, esencialmente– y de las circunstancias en que semejante instrumento presta sus servicios. Pero considera un síntoma pernicioso el silencio de Chomsky en torno a cuestiones como la de la función del contexto cultural en la adquisición del lenguaje. Reticencias como ésta, me parece a mí, responden a diferencias de matiz o acento, mejor que a discrepancias profundas. No es así ya en lo que toca a un punto distinto. Si Chomsky no avala una concepción instrumentalista del lenguaje, ello no se debe tanto a que piense que esa manera de ver las cosas es absolutamente errónea cuanto a que considera que hay formas más fructíferas de plantear *algunos* problemas relativos al lenguaje. Como se sabe, Chomsky ha defendido la propuesta de que una estructura cognitiva como el lenguaje ha de estudiarse de forma análoga a como lo haríamos con un órgano corporal. Una propuesta así insiste en la primacía metodológica del principio de la dependencia estructural: investigue Ud. la función del órgano, pero si no lo hace teniendo en cuenta su estructura, pocas cosas que no sean triviales podrá Ud. concluir, por expresarlo de forma muy grosera. El profesor Hierro no parece muy sensible a la fuerza del principio metodológico citado. Su crítica a Morris, por exigir éste que no deberíamos considerar lenguaje a un conjunto de signos que no se hallen relacionados entre sí según reglas que permitan ciertas combinaciones y que excluyan otras, así lo atestigua. Y a lo mismo apunta el que sostenga que "el único concepto de competencia [lingüística] que interesa es el de competencia de comunicación" (p. 63). Pero tanto en estos puntos específicos como en la cuestión más general, lo que hace falta es una prueba, o un argumento a favor, de que las cuestiones relativas a la estructura del órgano pueden resolverse a través del solo examen de sus funciones. Eso no parece nada fácil, y el mentalista y el innatista tienen en ello, puede argüirse, un asidero.

Con observaciones como las anteriores no quisiera dar la impresión de que el rechazo del programa psicolingüístico de Chomsky por parte del profesor Hierro se basa del todo en razones equivocadas. Esa impresión sería injusta. Muchas de sus críticas son completamente pertinentes, como las que afectan a las nociones barajadas de creatividad lingüística, a lo que el hablante-oyente *cognoce* ('*cognizes*') y al modelo "instantáneo" de adquisición del lenguaje. En mi opinión, estos y otros reparos deben aceptarse. Pero me parece igualmente muy pronto para concluir que el citado programa psicolingüístico es el producto de una metodología lisa y llanamente gratuita, dictada por un "mentalismo precientífico". (J. J. A.)

EL DIAGNOSTICO EN REUMATOLOGIA, por S. de Seze y A. Ryckewaert; 388 págs., 98 fig., 35 tablas; Toray-Masson, Barcelona 1980. En el prefacio de esta obra de Seze precisa su limitada aspiración: conducir al diagnóstico, a la clave de los procesos patológicos, que pueden englobarse dentro de la genérica denominación "reumáticos". No se analizan en ella, al menos detalladamente, los mecanismos que llevan hasta estos síndromes de origen articular, muscular, óseo, vascular y neurológico, no concretándose tampoco los criterios terapéuticos.

El libro consta de cinco partes. En la primera se examinan, en nueve capítulos, los reumatismos inflamatorios. Los tres primeros, redactados por los directores de la obra, abordan, sucesivamente, las poliartritis crónicas, las poliartritis agudas y las monoartritis agudas del adulto. El cuarto, redactado por Lequesneu, estudia las monoartritis subagudas y crónicas; el quinto y sexto, debidos a Welfling, atiende a los reumatismos intermitentes o recidivantes y a las artritis acompañadas de manifestaciones cutáneas; el séptimo, escrito por Kahn, se ocupa de las polialgias y de los estados dolorosos imprecisos; el octavo, elaborado por Tubiana y Goutalier, comprende un examen especial de las hemartrosis, y el último, referido a los cuerpos extraños intra-articulares, se debe a d'Angleau.

La segunda parte desarrolla en once capítulos un estudio topográfico de las artropatías. Ryckewaert, Caroit y Perez examinan las artropatías del hombro; de Seze y Maitre, las del codo; Grimet, Ryckewaert y de Seze, las de la mano; los dos últimamente citados y Perez se ocupan a continuación de la cadera crónica dolorosa del adulto; Gilmet y Ryckewaert pasan revista a las artropatías de la rodilla en el adulto; Dofmann se detiene en el estudio artroscópico de las afecciones dolorosas de la rodilla. En un capítulo ulterior Denis estudia la artropatía aislada del tobillo y las afecciones reumáticas del pie; Vitale se ocupa de la artropatía sacro-ilíaca, y Glimet y Djian de la artropatía temporo-maxilar, examinando éste las artropatías esterno-claviculares.

La tercera parte aborda, en diez capítulos, los síndromes dolorosos del raquis y de las raíces nerviosas, con un examen general de las radiculalgias elaborado por los directores de la obra; deformaciones vertebrales por Tellier y de Seze; cervicalgias posteriores en el adulto es redactado por Ryckewaert y Hubault; dorsalgias por de Seze, Guerin y Gurewitch; de Seze se ocupa de las lumbalgias; Ryckewaert y Hubault analizan las algias cervicobraqueales de origen radicular o pléxico. De Seze y Guerin abordan el diagnóstico de un dolor torácico, de Seze contempla luego los dolores de origen ciático, y, Vitale, la neuralgia crural. A continuación se investiga la inquietud nocturna de los miembros inferiores, examinando las molestias imprecisas de los mismos, de aparición nocturna con continuada movilidad de los miembros inferiores, integradas en un síndrome descrito en 1960 por Ekbom y que ya conoció Thomas Willis en el siglo XVIII denominándolo "restlessness"; interesante capítulo éste debido a Godlewski y Cherot.

La parte cuarta trata del diagnóstico de las osteopatías. Ryckewaert y Kuntz se ocupan de la osteoporosis; Bordier examina la osteomalacia, y, Kuntz, las condensaciones generalizadas del esqueleto; Cazalis estudia las hiperostosis múltiples del adulto; Lemaire, la enfermedad de Paget; Bard y de Seze, las hiperclaridades óseas localizadas. Al primero de éstos se debe el capítulo siguiente, diagnóstico de un aplastamiento vertebral único. El capítulo que cierra la sección, redactado por d'Anglejau, se refiere al diagnóstico de las fisuras óseas.

La última parte de esta obra aborda, en tres capítulos, algunos aspectos biológicos de los procesos patológicos que son objeto del libro. El primero y segundo contemplan, respectivamente, la actitud del médico ante una V.S. aumentada y ante la hipercalcemia; ambos han sido redactados por Kuhn. Finalmente, el último capítulo, firmado por Peltier, examina detalladamente la posición del médico ante alteraciones biológicas de significación precisa o incierta, factores reumatoides, anticuerpos antinucleares, paraproteínas, antiestreptolisina, etc. Sesenta y seis citas bibliográficas, integradas alfabéticamente, y un índice de materias con la misma ordenación completan el libro que acabamos de resumir.

Esta obra ha sido planteada con una doble finalidad: introducir al estudiante y al médico general en el diagnóstico de procesos dolorosos de origen diverso, localizados con variada topografía pero respondiendo siempre a una expresividad locomotora, y también como guía y medio de vocación para el especialista. Su claridad es evidente, nacida de la experiencia de los autores, y redactado con una expresión cuidada. Notoria es igualmente su limitación, no ya en cuanto a lo deliberadamente excluido, la etiopatogenia y los criterios de acción terapéutica, sino también en cuanto a un análisis más preciso de los síntomas derivados del examen fisiopatológico, tan enriquecedores para una racional valoración semiológica.

Su orientación estrictamente diagnóstica lleva a la reiteración, ya que determinados procesos son examinados en distintos lugares, unas veces con relación a su topografía y otras en cuanto a su evolución y a las manifestaciones acompañantes. Es evidente que una estimación nosográfica permitiría tras la visión de conjunto una precisión de las modalidades evolutivas, dentro de la fenomenología general del síndrome o entidad nosológica. Debemos señalar que ese no ha sido el propósito de los autores.

Las figuras, con un buen sentido esquemático, son magníficas y servirán a los que se inician para abrirse a la realidad del diagnóstico clínico-radiográfico. Muy adecuadas las tablas que sintetizan el desarrollo de diversos capítulos. La aportación de Kuhn sobre diagnóstico biológico es perfecta, tanto en lo concerniente a mecanismos de los fenómenos, sencilla y claramente expresados, cuanto en lo relacionado con su valoración aplicada al diagnóstico. Util es la aportación de Peltier, aunque integra un material más disperso y por ello menos accesible a una interpretación homogénea. Excelente la traducción realizada por José M.ª Poal Ballarin, esperada al proceder de un reumatólogo tan experto. (F. J. G.-C.)

ANTIQUE SCIENTIFIC INSTRUMENTS, por G. l'E Turner; Blandfort, Londres 1980. La técnica, como la ciencia de la que se nutre, constituye un continuum histórico. Desde las esquirlas y lascas del Paleolítico hasta la complejidad de una pastilla de microcomputador, la mano del hombre ha ido dominando la naturaleza para su mejor adaptación al medio y supervivencia. En no pocas ocasiones, el perfeccionamiento técnico ha hecho avanzar el conocimiento teórico, como ocurrió, primero, con la construcción del telescopio y, luego, con el de toda la gama de microscopios, por ceñirnos a la óptica.

La historiografía en torno a la técnica es bastante reciente, y, muy breve, la relativa a la habilidad española. Ausente casi por entero en la bibliografía extranjera, no ha merecido mejor suerte entre nuestros investigadores, reflejo, por otro lado, de una baja apreciación social de la "manufactura". Pero la inventiva doméstica ha sido bastante fértil y convendría que alguien la agrupase de un modo riguroso. Toledo, por ejemplo, destacó, en la Edad Media, en el campo de los útiles astronómicos. Que España sirvió de puente en la introducción en Occidente de importantes técnicas orientales (seda, papel) está bien documentado. El desarrollo de la náutica –ciencia, arte y técnica– a raíz del descubrimiento de América ha sido extensamente estudiado en *El arte de navegar en la España del Renacimiento* (Edit. Labor, 1979) por José María López Piñe o, quien también se ha ocupado de la ciencia y la técnica en el XVII, en particular del movimiento novator. Podríamos detenernos en el reinado de Carlos III, en la mente bulliciosa de Agustín de Betancourt. Hasta nuestros días: Monturiol, Torres Quevedo...

Se trata, empero, de figuras, de grupúsculos a lo más. La falta de museos de ciencia es un signo de ese distanciamiento entre lo académico –lo excepcional– y lo real –lo social–. De nuevo, las salvedades aisladas sólo sirven para confirmar ese desinterés público. El Museu de la Ciència de la Caja de Pensiones para la Vejez y de Ahorros de Cataluña y Baleares, en Barcelona, y el proyectado Museo Nacional de la Ciencia y la Tecnología, en Madrid, pueden paliar en parte esa necesidad de contacto directo entre el público y el producto científico recogido en ese buen hacer de los artesanos e inventores. Sería de desear que no se perdiera la ocasión entre burocracias y sinecuras y esas instituciones se convirtieran en verdaderas "casas de la ciencia", lugar de exhibición, experimentación y aprendizaje, al tiempo que ventana pública de la investigación en España y en el mundo.

Sirva este largo preámbulo para destacar la importancia del libro de Turner, cuyo destinatario es el coleccionista, aunque de su lectura pueden aprovecharse el historiador de la ciencia y los estudiosos de algunas ramas de ésta. Autor de varios libros sobre este campo, Turner ocupa un puesto directivo en el Museo de Historia de la Ciencia de Oxford, que guarda la mejor colección de instrumentos científicos del mundo.

Teniendo ante sí el potencial comprador de objetos científicos, de valor histórico-artístico, el autor simplifica al máximo las explicaciones de los instrumentos reseñados. (Los instrumentos tienen su lado de belleza, síntesis de la

ordenada perfección de la pieza acabada, de los materiales empleados, de la correcta aplicación de la teoría científica subyacente y de la habilidad del maestro que la construyó.) Sólo el especialista sabrá seguir el hilo sutil del progreso de la mecánica, la astronomía, la óptica o la medicina. La obra recuerda los catálogos de aficionados y coleccionistas: datos relativos al aparato, clases, función e importancia del mismo. La ficha técnica es muy completa dentro de su brevedad, está redactada con sumo rigor y en unos términos alcanzables por los no expertos. Otra limitación: sólo se contemplan objetos que el coleccionista puede adquirir, es decir, que puedan haberse conservado hasta nuestros días.

Se agrupan los instrumentos en tres grandes categorías: matemáticos, ópticos y físicos, según la autodefinición que de su profesión dio Dudley Adams en el siglo XVIII: "Mathematical, Optical and Philosophical Instrument Maker". Los instrumentos matemáticos se denominan así en razón de su finalidad: medir el tiempo, la longitud y la masa. Astrolabios, relojes (de sol y arena), cuadrantes, planetarios, bitácoras, balanzas, calculadoras, pesas, medidas de capacidad...

De reconocido valor para el avance del saber fue el desarrollo de los instrumentos ópticos. Las propiedades de los espejos que descubriéronse en la Antigüedad, y cuyo exponente más familir puede ser Arquímedes, se aprovecharon en la Edad Media para la instrucción del pueblo –cristaleras historiadas–, entre otros usos. Turner opina que el retraso registrado en el desarrollo de los instrumentos ópticos se debió, en parte, a la pobre calidad de los cristales, lo que explica que los telescopios del XVIII fueran de metal pulimentado. Microscopios, telescopios, cámaras, caleidoscopios... concitaron la destreza de los artesanos y un mayor conocimiento de las leyes de la reflexión, refracción, dispersión de la luz por un prisma, etcétera.

Los instrumentos físicos se idearon para producir y demostrar los efectos de ciertas leyes de la mecánica, el magnetismo, la neumática, la hidrostática, la electricidad, el calor y la meteorología. Sin olvidarse los que incidieron en el campo médico y quirúrgico: catéteres, estetoscopio (que revolucionó la práctica y el diagnóstico médico en el siglo pasado). Una bibliografía sumaria, oportunos consejos para el coleccionista poco avisado y un elenco de los principales museos cierran esta obra cuyo aparato ilustrativo es, sencillamente, único. (L. A.)

Bibliografía

Los lectores interesados en una mayor profundización de los temas expuestos pueden consultar los trabajos siguientes:

LA FABULA

A New Look at the Wisdom of Sumer and "Akkad". E. Gordon en *Bibl. Orientalis,* vol. 17, págs. 122-152; 1960.

La Fable Antique. I. M. Nøjgaard; Copenhague, 1964; II. id. 1967.

Das Lateinische Tierepos. F. P. Knapp; Darmstadt, 1979.

Historia de la Fabula Greco-Latina. I. Introduccion y De los Origenes a la Edad Helenistica. F. R. Adrados; Madrid, 1979.

GALILEO OBSERVA NEPTUNO

Mutual Occultations of Planets: 1557 to 2230. Steven C. Albers en *Sky and Telescope,* vol. 57, n.° 3, págs. 220-222; marzo, 1979.

Galileo and Satellite Prediction. Stillman Drake en *Journal for the History of Astronomy,* vol. 10, n.° 28, 2.ª parte, págs. 75-95; junio, 1979.

Galileo's Observations of Neptune. Charles T. Kowal y Stillman Drake en *Nature,* vol. 287, n.° 5780, págs. 311-313; 25 de septiembre de 1980.

ASIMETRIA COSMICA ENTRE MATERIA Y ANTIMATERIA

Observational Tests of Antimatter Cosmologies. Gary Steigman en *Annual Review of Astronomy and Astrophysics,* vol. 14, págs. 339-372; 1976.

Matter-Antimatter Accounting, Thermodynamics, and Black-Hole Radiation. D. Toussaint, S. B. Treiman, Frank Wilczek y A. Zee en *Physical Review D,* vol. 19, n.° 4, págs. 1036-1045; 15 de febrero de 1979.

PLASMIDOS

Extrachromosomal Inheritance in Bacteria. R. P. Novick en *Bacteriological Review,* vol. 33, págs. 210-263; 1969.

Replication of a Bacterial Episome under Relaxed Control. R. Rownd en *Journal of Molecular Biology,* vol. 44, n.° 3, págs. 387-402; 28 de septiembre de 1969.

Infectious Multiple Drug Resistance. Stanley Falkow. Academic Press, 1975.

On Plasmid Incompatibility. Richard P. Novick y F. C. Hoppensteadt en *Plasmid,* vol. 1, n.° 4, págs. 421-434; septiembre, 1978.

Control of Plasmid R1 Replication: Functions Involved in Replication, Copy Number Control, Incompatibility, and Switch-off of Replication. Søren Molin y Kurt Nordström en *Journal of Bacteriology,* vol. 141, n.° 1, págs. 111-120; enero, 1980.

EL HOMBRE PRIMITIVO EN EL ASIA CENTRAL SOVIETICA

Karatau I-Drevneyshiy paleoliticheskiy pamyatnik v lessakh Sredney Azii (Karatau I-la Estacion Paleolitica Mas Antigua de los Loess del Asia Central). A. A. Lazarenko y V. A. Ranov en *Byulleten' Kommissii po Izucheniyu Chetvertichnogo Perioda,* n.° 47, págs. 45-47; 1977.

Nakhodki Paleolita v drevnikh pogrebennykh pochvakh yuzhnogo Tadzhikistana i ikh geologicheskaya pozitisiya (Hallazgos Paleoliticos en Suelos Antiguos Enterrados del Tadjikistan Meridional y su Posicion Geologica). A. E. Dodonov, V. A. Ranov y A. V. Pen'kov en *Byulleten' Kommissii po Izucheninyu Chetvertichnogo Perioda,* n.° 48, págs. 86-102; 1978.

Toward a New Outline of the Soviet Central Asian Paleolithic. V. A. Ranov y R. S. Davis en *Current Anthropology,* vol. 20, n.° 2, págs. 249-270; junio, 1979.

LA CONDUCTA ADICTIVA

Methadone Plasma Level: Sustained by a Reservoir of Drug in Tissue. Vincent P. Dole y Mary Jeanne Kreek en *Proceedings of the National Academy of Sciences of the United States of America,* vol. 70, nº 1, pág. 10, enero, 1973.

Drug Interactions with Methadone. Mary Jeanne Kreek, Chai Leng Gutjhar, Jane W. Garfield, David V. Bowen y Frank H. Field en *Annals of the New York Academy of Sciences,* vol. 261, págs. 350-371; 10 de diciembre de 1976.

Methadone Treatment in Narcotic Addiction: Program Management, Findings, and Prospects for the Future. Robert G. Newman en colaboración con Margot S. Cates. Academic Press, 1977.

EL SISTEMA DE HUIDA DE LA CUCARACHA

Responses of Giant Interneurons of the Cockroach *Periplaneta americana* to Wind Puffs of Different Directions and Velocities. J. Westin, J. Langberg y J. Camhi en *Journal of Comparative Physiology,* vol. 121, págs. 307-324; 1977.

The Escape Behavior of the Cockroach *Periplaneta americana,* II: Detection of Natural Predators by Air Displacement. J. Camhi, W. Tom y S. Volman en *Journal of Comparative Physiology,* vol. 128, págs. 203-212; 1978.

FORMACION DE IMAGENES EN LOS ESPEJOS

The Ambidextrous Universe: Left, Right, and the Fall of Parity. Martin Gardner. New American Library, 1969.

The Left Hand of the Electron. Isaac Asimov. Doubleday & Co., Inc., 1972.

Geometry. Harold R. Jacobs. W. H. Freeman and Company, 1974.

JUEGOS MATEMATICOS

Let Us Now Praise Prime Numbers. Helen Spalding en *An Anthology of Modern Verse 1940-1960,* dirigido por Elizabeth Jennings. Methuen & Co. Ltd, 1961.

Searching for the 27th Mersenne Prime. David Slowinski en *Journal of Recreational Mathematics,* vol. 11, n.° 4, págs. 458-461; 1978-79.

A Search for Large Twin Prime Pairs. R. E. Crandall y M. A. Penk en *Mathematics of Computation,* vol. 33, n.° 145, págs. 383-388; enero, 1979.

TALLER Y LABORATORIO

An Effect of Polarized Sky Light. S. G. Cornford en *Weather,* vol. 23, n.° 1, págs. 39; enero, 1968.

Sundials: Their Theory and Construction. Albert E. Waugh. Dover Publications, Inc., 1973.

Seguiremos explorando los campos del conocimiento



HISTORIA PARALELA DEL TELEFONO, por David A. Hounshell
Cuando Alexander Graham Bell desarrollaba el teléfono, lo mismo hacía Elisha Gray. Bell consiguió la patente, pero el hecho no deja de ser un instructivo ejemplo de invención simultánea.

VIROIDES, por T. O. Diener
Se trata de los agentes más pequeños de enfermedades infecciosas que se conoce: son filamentos cortos de ARN. Causan distintas enfermedades vegetales y posiblemente estén implicados en enigmáticas dolencias del hombre y otros animales.

RACCHI (PERU), UN ENIGMA ARQUEOLOGICO, por Manuel Ballesteros Gaibrois
Magnitud, grandeza y audacia de la construcción de un complejo inca.

EL CORAZON ARTIFICIAL TOTAL, por Robert K. Jarvik
Los sustitutos mecánicos del corazón natural siguen perfeccionándose. Con el tiempo estarán listos para ser implantados en los pacientes que lo necesiten.

LA GALAXIA DE ANDROMEDA, por Paul W. Hodge
Esta gran galaxia espiral, la más próxima a la nuestra, se ha constituido en verdadero laboratorio para el estudio de la evolución de las estrellas y las galaxias.

LOS RECURSOS GENETICOS SILVESTRES DEL TRIGO, por Moshe Feldman y Ernest R. Sears
Las prácticas modernas de mejora han reducido la variabilidad genética de los trigos cultivados. La mejor esperanza para futuras mejoras de esta planta se funda en la explotación del abundante acervo genético de las especies silvestres afines.

EL PROBLEMA MENTE-CUERPO, por Jerry A. Fodor
¿Pueden sufrir las máquinas, albergar esperanzas los marcianos y pensar los espíritus separados de su cuerpo? De acuerdo con el funcionalismo moderno, cabe la posibilidad lógica de que así ocurra.

GELES, por Toyoichi Tanaka
Están formados en su mayor parte por un fluido al que da forma una red de filamentos poliméricos. Un equilibrio de fuerzas mantiene su estructura; si se altera, siquiera infinitesimalmente, puede producirse una transición de fase y la descomposición del gel.

